

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 23:04:53 ; Search time 4244.29 Seconds  
(without alignments)  
10279.384 Million cell updates/sec

Title: US-10-668-047-1

Perfect score: 1461

Sequence: 1 gtgcattccgacgcagcac.....ttccggaatcgctcgtga 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	47.4	3.2	1546	28	BZ568822
2	45.8	3.1	794	28	BZ560168
3	44.8	3.1	922	9	AL571419
4	43.6	3.0	925	29	CNS0091P
					AL053013 Drosophil

C 5	42.6	2.9	690	13	CA136496
C 6	41.6	2.8	718	13	CA131027
C 7	41.4	2.8	1101	29	CNS017SY
8	41.4	2.8	1201	13	EX381961
9	41.2	2.8	508	12	BM727423
C 10	41.2	2.8	770	28	BZ563177
C 11	41.2	2.8	1347	28	BZ574354
C 12	41	2.8	579	12	BM712537
C 13	41	2.8	596	12	BJ469381
C 14	41	2.8	661	12	BJ472412
C 15	40.8	2.8	1201	13	EX356664
C 16	40.2	2.8	721	14	CD494071
C 17	40.2	2.8	1176	14	CD502787
C 18	39.8	2.7	802	12	BI183587
C 19	39.8	2.7	885	29	CG369642
20	39.8	2.7	885	13	EX425603
21	39.8	2.7	958	13	CA151619
22	39.6	2.7	514	10	BF594332
23	39.6	2.7	901	29	CG353488
24	39.6	2.7	921	29	CG446568
25	39.4	2.7	459	12	BM696019
26	39.4	2.7	1029	9	AL528764
C 27	39.2	2.7	625	29	CC711484
C 28	39.2	2.7	806	28	BZ402559
C 29	39.2	2.7	809	29	CG627722
C 30	39.2	2.7	918	29	CG245606
C 31	39.2	2.7	971	29	CG240394
C 32	39.2	2.7	1198	12	BI957046
C 33	39	2.7	949	29	CC930329
C 34	38.8	2.7	391	14	CB879070
C 35	38.8	2.7	430	14	CD036955
C 36	38.8	2.7	507	12	BM075546
C 37	38.8	2.7	524	14	CD036685
C 38	38.8	2.7	534	12	BM335322
C 39	38.8	2.7	542	12	BM335465
C 40	38.8	2.7	572	12	BM335137
C 41	38.8	2.7	574	12	BJ088126
C 42	38.8	2.7	557	12	BM349453
C 43	38.8	2.7	729	28	BZ528408
C 44	38.8	2.7	787	29	CG360243
C 45	38.8	2.7	837	10	BF621278

#### ALIGNMENTS

RESULT 1	BZ568822	1546 bp	DNA	linear	GSS 17-DEC-2002
LOCUS	pac2-164_8106.y2	pac2-164	Pseudomonas aeruginosa	genomic clone	
DEFINITION	pac2-164_8106.y2	pac2-164	Pseudomonas aeruginosa	genomic clone	
ACCESSION	BZ568822	GI:27202436			
VERSION	BZ568822.1	GI:27202436			
KEYWORDS	GSS.				
SOURCE	Pseudomonas aeruginosa				
ORGANISM	Pseudomonas aeruginosa				
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.				
AUTHORS	Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M., Burns, J.L., Kaul, R. and Olsen, M.V.				
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library				
JOURNAL	J. Bacteriol. (2002) In press				
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: shotgun.				
FEATURES	Location/Qualifiers				

```

source
1. .1546
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_8106"
/cloned_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

ORIGIN
Query Match 3.2%; Score 47.4; DB 28; Length 1546;
Best Local Similarity 53.2%; Pred. No. 0.46;
Matches 123; Conservative 0; Mismatches 106; Indels 2; Gaps 1;

QY 149 GAACGGTGGCGGTGCTGACCCAGGAGCTGGTGTGGGTCTCTGACGGCTGCCCTCGTCG 208
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
389 GAGCTGATCAGATCTCTCAAGCAGGAGAGGCTGCGGGTCCGAAGACGCTGGTCTATTTCG 448
QY 209 ATCGGTGTCATCTGAACGCCGCCGACCTTCGGGTCCACGTAGTTGCCGTGGAAACCGACC 268
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
449 GTCTACACCGGCAGCGCGACACACGCTCGGCTTGAAGTGTGTC--TGGAGCAGGAAG 506
QY 269 CTTTGTGCTGCTTACTCTCGGCGCACCTCGAGGAATGTGGAAAGCTTACGGCATCT 328
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
507 GCTTCAAGGTGGCGGTGCTCGGCGAAGCTGGATGCTCCCGCGCGAGGACTGGATCG 566
QY 329 CCTACGACCTAGTCGAGGGCGACTATTGCTTACCAAGGGGCCAAGCTGG 379
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
567 CCGAGCAGTTGGACCGCGGATCGACGTCTGATCACGAACCCCGAGCTGG 617

RESULT 2
BZ560168 794 bp DNA linear GSS 17-DEC-2002
LOCUS pacs2-164_2210.xl pacs2-164 Pseudomonas aeruginosa genomic clone
DEFINITION pacs2-164_2210, genomic survey sequence.
ACCESSION BZ560168
VERSION BZ560168.1 GI:27177735
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 794)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .794
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164_2210"
/cloned_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."

FEATURES
source
1. .794
Query Match 3.1%; Score 45.8; DB 28; Length 794;
Best Local Similarity 52.8%; Pred. No. 0.86;
Matches 122; Conservative 0; Mismatches 107; Indels 2; Gaps 1;

```

```

QY 680 AAGAGATGTCATCGTTCGCTTTCAGTGTAGGCC 713
    | | | | | | | | | | | | | | | | | | | |
Db 791 AGGGAATGCMRMATTTMAAVMRMGWGGMM 824

RESULT 4
CNS0091P
LOCUS
DEFINITION
  CNS0091P 925 bp DNA linear GSS 03-JUN-1999
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL053013
  AL053013.1 GI:4934461
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 925)
REFERENCE
  Genoscope.
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr
  - Web : www.genoscope.cns.fr)
COMMENT
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazutoyo Osogawa and
  Aaron Mammosier in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPCI-98 and was constructed by partial,
  EcoRI digestion of Drosophila DNA provided by the BDGP from the
  isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
  P1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
  Location/Qualifiers
    1..925
      /organism="Drosophila melanogaster"
      /mol_type="genomic DNA"
      /db_xref="taxon:7227"
      /clone="BACR19D16"
      /clone_lib="RPCI-98"
      /note="end : TET3"

ORIGIN
Query Match      3.0%; Score 43.6; DB 29; Length 925;
Best Local Similarity 13.2%; Pred. No. 3.6;
Matches 47; Conservative 157; Mismatches 152; Indels 0; Gaps 0;

QY 28 CGCGCCAGACTGCTCTCGGCAAACTGGACCCCACTACTCAAGCGGTGTAGCGAGTTC 87
    | | | | | | | | | | | | | | | | | | | |
Db 570 BSCGCCSGSCSSSSGCCBCCCCCSCSYCCSSBSKSTBSKSCCCSKSVCGT 629

QY 88 TTCACTCCCATGAAGCCGCCAGCTGATGGCTTCAATGCTTCGGTTCATGATCTCCGC 147
    | | | | | | | | | | | | | | | | | | | |
Db 630 SCSSSSSSSSSSSTSSSTSSSTSSKSSSSSSSSSSSYTTSKTSASGSGSWAGGGS 689

QY 148 GGAAACGGTGGCGGTGTCGACCCAGAGCTGGGTGTCGGGTCTCTGACCGCTGCCCTCGNC 207
    | | | | | | | | | | | | | | | | | | | |
Db 690 GGTGTTSSSSSSSTSSSVSSGSKSTBSBSGSSSGSSSSSTSSBBSCTSTSSSS 749

QY 208 GATCGCTGATACTGAACGCCGCCAGCTGTGGGTCCACGTAGTTCCCGTGGAAACCCGAC 267
    | | | | | | | | | | | | | | | | | | | |
Db 750 SSSYSSTSCCTCCCSYSSSTSSSSSTSWGSTSGSSSSSVGTRSSSDSTTCSCCC 809

QY 268 CCCTTTGTGTCCTTACTCGCGCCACACCTGGAGGAATGCGGAACGCTTAGCGATC 327
    | | | | | | | | | | | | | | | | | | | |
Db 810 YMCTCCTBYMBCYTSCGGSSSSSGKGGVTKCGCGGSSSTNGMBGTSSACSSSSSS 869

```



[illegible]

Db 147 CAGCCCGGTGTCCCGTATCTCGCGCATCGCTCGGTCCGCGCAGCGTCCATCCAGGCG 206  
 QY 230 CCGACGTTGGTCCACGATGTCGCGTGAACCGACCCCTTTGTCTGCTTACCTGC 289  
 Db 207 TGCAGGTGCGCGTGGCGTGAAGTCCGCGAGCGATCCGAATCCGCTCGGTACCGCC 266  
 QY 290 GCGCCACCTCGAGGAATGTCCGAACGCTTACGGCATCTCTACGACCTAGTCGAGGCG 349  
 Db 267 CGCCCGTCAAAACCCAGGTGCGACGTGNNCGAGATATCCGCCACCTCTGTCGCG 326  
 QY 350 ACTA 353  
 Db 327 CCTA 330

RESULT 10  
 BZ563177/c  
 LOCUS  
 DEFINITION pacs2-164\_4126.y2 pacs2-164 Pseudomonas aeruginosa genomic clone  
 pacs2-164\_4126, genomic survey sequence.

ACCESSION BZ563177.1 GI:27186541  
 VERSION  
 KEYWORDS  
 SOURCE GSS.

ORGANISM Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 770)  
 AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,  
 Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..770

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="2-164"

/db\_xref="taxon:287"

/clone="pacs2-164\_4126"

/clone\_lib="pacs2-164"

/note="clinical isolate 2-164 Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 770;  
 Best Local Similarity 50.5%; Pred. No. 14;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 129 TCGGTCGATGATCTCCGCGAAGTGGCGGTCTCGACCCGAGGCTGGTGTGGGTC 188  
 Db 341 TCAGCGCAGGTCTCTCCGCGAGGCGCTGAATCCATGCCACGCGAGGCTTCGCCGATCA 282

QY 189 TCTGACCGTGCCTCGTGCATCGGTGATGAAACGCCCGGAGCTTGGCGTCCACGT 248  
 Db 281 TCTGCTGACGCCCGCGCCGATGATGTCACGCCGAGCACCCTGCTCGCGGATCG 222

QY 249 AGTTCGCGTGAACCGACCCCTTGTGTCGCTTACCTGCGCGCCACCTCGAGGAATG 308  
 Db 221 AGAGGATCTTGAAGAGCTTCTGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 162

QY 309 TCGGACCGTTAGGCGAT 326

Db 161 ACGGGAAGCGTCCGACCT 144

RESULT 11

BZ574354/c

LOCUS

DEFINITION msh2\_363.y2 msh pseudomonas aeruginosa genomic clone msh2\_363,

genomic survey sequence.

ACCESSION BZ574354

VERSION BZ574354.1 GI:27209415

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..1347

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="MSH"

/db\_xref="taxon:287"

/clone="msh2\_363"

/clone\_lib="msh"

/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 1347;  
 Best Local Similarity 50.5%; Pred. No. 19;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 129 TCGGTCGATGATCTCCGCGAAGTGGCGGTCTCGACCCGAGGCTGGTGTGGGTC 188

Db 653 TCAGCACGAGGTCTCTCCGCGAGGCGCTGAATCCATGCCACGCGAGGCTTCGCCGATCA 594

QY 189 TCTGACCGTGCCTCGTGCATCGGTGATGAAACGCCCGGAGCTTGGCGTCCACGT 248

Db 593 TCTGCTGACGCCCGCGCGGATCATGTGACGCCGAGCACCTGGTGTGCTCGCGGATCG 534

QY 249 AGTTCGCGTGAACCGACCCCTTGTGCTGCTTACCTGCGCGCACCTCGAGGAATG 308

Db 533 AGAGATCTTGAAGAGCTTCTGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 474

QY 309 TCGGAACGCTTACGCGAT 326

Db 473 ACGGGAAGCGTCCGACCT 456

RESULT 12

BM712537

LOCUS

DEFINITION UI-E-EJO-ahh-k-14-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone

UI-E-EJO-ahh-k-14-0-UI 5', mRNA sequence.

ACCESSION BM712537.1 GI:19025795

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 579)

BZ574354 1347 bp DNA linear GSS 17-DEC-2002  
 msh2\_363.y2 msh pseudomonas aeruginosa genomic clone msh2\_363,  
 genomic survey sequence.

ACCESSION BZ574354

VERSION BZ574354.1 GI:27209415

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

COMMENT Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1..1347

/organism="Pseudomonas aeruginosa"

/mol\_type="genomic DNA"

/strain="MSH"

/db\_xref="taxon:287"

/clone="msh2\_363"

/clone\_lib="msh"

/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 2.8%; Score 41.2; DB 28; Length 1347;  
 Best Local Similarity 50.5%; Pred. No. 19;  
 Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 129 TCGGTCGATGATCTCCGCGAAGTGGCGGTCTCGACCCGAGGCTGGTGTGGGTC 188

Db 653 TCAGCACGAGGTCTCTCCGCGAGGCGCTGAATCCATGCCACGCGAGGCTTCGCCGATCA 594

QY 189 TCTGACCGTGCCTCGTGCATCGGTGATGAAACGCCCGGAGCTTGGCGTCCACGT 248

Db 593 TCTGCTGACGCCCGCGCGGATCATGTGACGCCGAGCACCTGGTGTGCTCGCGGATCG 534

QY 249 AGTTCGCGTGAACCGACCCCTTGTGCTGCTTACCTGCGCGCACCTCGAGGAATG 308

Db 533 AGAGATCTTGAAGAGCTTCTGCTCTCGTGGTGTATCTTCGCCCGGCTGTTGGCGCTGA 474

QY 309 TCGGAACGCTTACGCGAT 326

Db 473 ACGGGAAGCGTCCGACCT 456

BM712537 579 bp mRNA linear EST 28-FEB-2002  
 UI-E-EJO-ahh-k-14-0-UI.r1 UI-E-EJO Homo sapiens cDNA clone  
 UI-E-EJO-ahh-k-14-0-UI 5', mRNA sequence.

ACCESSION BM712537.1 GI:19025795

VERSION EST.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 579)

**AUTHORS** Bonaldo,M.F., Lennon,G. and Soares,M.B.  
**TITLE** Normalization and subtraction: two approaches to facilitate gene discovery  
**JOURNAL** Genome Res. 6 (9), 791-806 (1996)  
**MEDLINE** 97044477  
**PUBMED** 8889548  
**COMMENT** Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com).  
 Seq primer: M13 Reverse.

**FEATURES** Location/Qualifiers  
 source 1..579

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="UI-E-EJ0-ahh-k-14-0-UI"  
 /tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and Choroid"  
 /dev\_stage="fetal and adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /clone\_lib="UI-E-EJ0"  
 /note="Organ: eye; Vector: pUT3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ0 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pUT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAAGT; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

## ORIGIN

Query Match 2.8%; Score 41; DB 12; Length 579;  
 Best Local Similarity 51.4%; Pred. No. 14;  
 Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 169 CCAGGAGCTGGTCCGGTCTCTGACCGCTGCCCTCGTCGATCGGTCGATCTGAACGC 228  
 |||||  
 Db 146 CCAGGCCCGGTGTCGGTATCTCGGGATCCGCTGGTCCGGCAGCGATCATCCAGGC 205  
 |||||  
 QY 229 CCGGACGTTGCGGTCCACGTAGTTCGGTGGAAACCGAACCCCTTTGTCGTCCTTACCTG 288  
 |||||  
 Db 206 GTGCAGGTCCGGTGGCGTGGTGAAGTCCGAGCGATCCGAATCCGGTCCGGTACCCG 265  
 |||||  
 QY 289 CGCGCCACCTCGAGGATGTCGGAAGCTTACGGCATCTCCTACGACCTAGTCGAGGGC 348  
 |||||  
 Db 266 CGCGCCGCTCAAAACCCAGGTCGCGACGTGGTCGAGATATCCGCCACCACTCGTCGCG 325  
 |||||  
 QY 349 GACTA 353  
 |||||  
 Db 326 GCCTA 330  
 |||||

RESULT 13

BJ469381/c  
**LOCUS**  
**DEFINITION**

BJ469381 596 bp mRNA linear EST 23-MAY-2002  
 heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA clone baal20k01 5', mRNA sequence.

**ACCESSION** BJ469381  
**VERSION** BJ469381.1 GI:21147884  
**KEYWORDS** EST.  
**SOURCE** Hordeum vulgare subsp. vulgare  
**ORGANISM** Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.

1 (bases 1 to 596)

Sato,K., Saisho,D. and Takeda,K.  
 Barley EST sequencing project in NIG and Okayama Univ  
 Unpublished (2002)

**JOURNAL** Contact: Tadasu Shin-i

**COMMENT** Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

**FEATURES** Location/Qualifiers

source 1..596

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="Haruna Nijo"

/sub\_species="vulgare"

/db\_xref="taxon:112509"

/clone="baal20k01"

/tissue\_type="top three leaves"

/dev\_stage="adult, heading stage"

/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna Nijo adult, heading stage top three leaves"

## ORIGIN

Query Match 2.8%; Score 41; DB 12; Length 596;  
 Best Local Similarity 52.7%; Pred. No. 14;  
 Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 124 ATCTTCGGGTGCGATGATCTCCCGGAACGGTGGGGTCTCCAGCCAGAGTGGTGC 183  
 |||||

Db 197 ATCTTCGGGGCCCTCGTGTGCGAGAGGGGTACTCGTCGATGACCCAGTTGGTGGTC 138  
 |||||

QY 184 GGGTCTCTGACCGCTGCCCTCGTCGATCGGCTGCATCTGAAACGCCCGACGTTGGGTC 243  
 |||||

Db 137 GCCTTTGGGAGCGCGCGCGGTGGAAGACGACGCTCTTCTTCATCCGACGAGGGGCC 78  
 |||||

QY 244 CAGCTAGTTGCGGTGGAACCGGACCCCTTTGTCGTCGCTTACCTGCGCG 292  
 |||||

Db 77 GCGGTGGGACACCTCGCGGTCTTTCGCGGGTCTTCCAGTAGCGCGCG 29  
 |||||

RESULT 14

BJ472412/c

**LOCUS**

**DEFINITION**

BJ472412 661 bp mRNA linear EST 23-MAY-2002  
 heading stage top three leaves Hordeum vulgare subsp. vulgare cDNA clone baal34e02 5', mRNA sequence.

**ACCESSION** BJ472412

**VERSION** BJ472412.1 GI:21150915

**KEYWORDS** EST.

**SOURCE** Hordeum vulgare subsp. vulgare

**ORGANISM** Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Poideae; Triticeae; Hordeum.

1 (bases 1 to 661)

Sato,K., Saisho,D. and Takeda,K.

Barley EST sequencing project in NIG and Okayama Univ

Unpublished (2002)

COMMENT Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
1. .661  
/organism="Hordium vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna Nijo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baal34e02"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna  
Nijo adult, heading stage top three leaves"

FEATURES  
source  
1. .661  
/organism="Hordium vulgare subsp. vulgare"  
/mol\_type="mRNA"  
/cultivar="Haruna Nijo"  
/sub\_species="vulgare"  
/db\_xref="taxon:112509"  
/clone="baal34e02"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, cv. Haruna  
Nijo adult, heading stage top three leaves"

ORIGIN  
Query Match 2.8%; Score 41; DB 12; Length 661;  
Best Local Similarity 52.7%; Pred. No. 15;  
Matches 89; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 124 ATGCTTCGGTGCATGATCTCGCGAAGCGTGGCGGTCTCGACCCAGGAGCTGGTCTC 183  
Db 196 ATCTTGGGGCGCGTCTGTCAGAGAGCGGTACTGTCGTCATGACCCAGTGGTGGTCTC 137  
QY 184 GGGTCTCTGACCGTCCCTCGTCGATCGGTGCTGATGACGACCCCGACGTTCCGGTCTC 243  
Db 136 GCCCTTGGAGCGCGCGCGGTGAGAGAGCGTCTTCTTCATCCGACGAGCGCGCC 77  
QY 244 CAGCTAGTTCGGTGAACCGACCCCTTGTGCGCTTACCTACCTGGCG 292  
Db 76 GCGGTGGACACCTCGCGTCTTTCGCGGGTCTTCCAGTAGCCGCGCG 28

ORIGIN  
Query Match 2.8%; Score 40.8; DB 13; Length 1201;  
Best Local Similarity 11.1%; Pred. No. 23;  
Matches 37; Conservative 160; Mismatches 135; Indels 0; Gaps 0;  
QY 65 CTCAGCGGTCTAGGCGAGTTCCTCACTCCCATGAAGCGCCGACGCTGATGGCTTCAA 124  
Db 774 SYBSBSBTTSSCCTSSBSSTTSBMTSSBTTSSSSSSSSSBTTSTTSKSBTT 833  
QY 125 TGCTTGGGTGCATGATCTCCGCGAAGCGTGGGTCTCTGACCCAGGAGCTGGTGTGCG 184  
Db 834 TSBBTSSSSSTSTWTBTCTSTTNTTSSYSBSSTSTBTSSTTSBSSBTTTTSBT 893  
QY 185 GGTCTCTGACCGTGCCTCGTCTGATCGGTGCTGATGAAACGCCCGACGTTGCGGTCC 244  
Db 894 STBTSSSSSTTSSSSSSSSSTTSSSBTTSSSBTTSSSSSSSTSSSSSSSS 953  
QY 245 ACCTAGTTGCGGTGGAACCGACCCCTTGTGCGCTTACTGCGCGCCACCTGGAGG 304  
Db 954 SSSBTTSSKSSSSSSSSSSSSSTTBSSTTTTTSBTTSSBTTSSSTSSSTSS 1013  
QY 305 AATGTCGGAACGCTTACGGCATCTCTACGACCTAGTCTGAGGCGGACATTGCTTAACC 364  
Db 1014 BSSNTSSSSSSSSSTBTSSSNKSSSSSSSSSSSSSSSBTTTTSBTTT 1073  
QY 365 AAGGGCGCAAGCTGGATGCGCCGTTTCGATCTT 396  
Db 1074 BSSBSBTTTSSBSBSSSTTSBTTSSSTTTT 1105

Search completed: October 2, 2004, 05:12:39  
Job time : 4258.29 secs

RESULT 15  
BX356664  
LOCUS  
DEFINITION BX356664 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
clone CS0D1015YB03 3-PRIME, mRNA sequence.  
ACCESSION BX356664  
VERSION BX356664.1 GI:30378083  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1201)  
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CS0D1015CA02NP1.

FEATURES  
source  
1. .1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D1015YB03"  
/tissue\_type="PLACENTA COT 25-NORMALIZED"  
/clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:25:51 ; Search time 5949.09 Seconds  
(without alignments)  
10644.353 Million cell updates/sec

Title: US-10-668-047-1

Perfect score: 1461

Sequence: 1 gtgacccgatcgccagcac.....ttccggaatctgctgtga 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_ov.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_mam.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	72.2	4.9	3889	1	PROIRM	K02081 P.stuartii
C 2	59.4	4.1	343250	1	AF0003594	AF0003594 Nostoc sp
C 3	51.2	3.5	1600	1	AF146609	AF146609 Aeromonas
C 4	49.6	3.4	2000	6	AX655393	AX655393 Sequence
C 5	45	3.1	2892	1	BACHSMR	L01541 Bacillus su
C 6	45	3.1	10929	1	U67554	U67554 Methanococ
C 7	45	3.1	110000	6	AR271569_04	Continuation (5 of
C 8	44.6	3.1	5039	1	YEN14030	Continuation (5 of
C 9	44.2	3.0	1164	8	AK108161	AK108161 Oryza sat
C 10	44.2	3.0	146585	8	CNS08CAM	AK1831809 Oryza sat
C 11	43	2.9	300900	1	AP005939	AP005939 Bradyrhiz
C 12	42.8	2.9	17786	1	RLHUPGENE	AP005939 Bradyrhiz
C 13	42.8	2.9	21376	12	AF246703	AF246703 Synthetic
C 14	42.8	2.9	175376	2	AC108223	AC108223 Oryza sat
C 15	42.4	2.9	125020	9	AF429315	AF429315 Homo sapi
C 16	42.2	2.9	125020	9	AF429315	AF429315 Homo sapi
C 17	42	2.9	12337	1	AE011861	AE011861 Xanthomon
C 18	41.6	2.8	12046	1	AE005017	AE005017 Halobacte
C 19	41.2	2.8	11735	1	AE004896	AE004896 Pseudomon
C 20	41.2	2.8	12766	1	RLHUPREG	236981 R.leguminos
C 21	41.2	2.8	113193	1	AF357202	AF357202 Streptomy
C 22	41.2	2.8	113193	6	AX703543	AX703543 Sequence
C 23	41.2	2.8	132544	1	AF521085	AF521085 Streptomy
C 24	41.2	2.8	299925	1	AP005045	AP005045 Streptomy
C 25	41	2.8	299950	1	AP005963	AP005963 Bradyrhiz
C 26	40.8	2.8	302178	1	AE016918	AE016918 Chromobac
C 27	40.6	2.8	13993	1	AE001890	AE001890 Deinococc
C 28	40	2.7	302578	1	AE017208	AE017208 Geobacter
C 29	39.8	2.7	10393	1	AE011651	AE011651 Xanthomon
C 30	39.8	2.7	13325	1	AE005016	AE005016 Halobacte
C 31	39.8	2.7	219986	2	AC113026	AC113026 Mus muscu
C 32	39.8	2.7	299925	1	AP005045	AP005045 Streptomy
C 33	39.6	2.7	2000	6	AX655393	AX655393 Sequence
C 34	39.6	2.7	5364	1	SGKSACPG	X77865 S.griseus g
C 35	39.6	2.7	11847	1	AE001918	AE001918 Deinococc
C 36	39.6	2.7	300140	1	AP005956	AP005956 Bradyrhiz
C 37	39.4	2.7	12208	1	SGPREPG	X80774 Streptomyce
C 38	39.4	2.7	302101	1	AE016784	AE016784 Pseudomon
C 39	39.2	2.7	131676	9	AC139749	AC139749 Homo sapi
C 40	39.2	2.7	160766	2	AC137686	AC137686 Homo sapi
C 41	39	2.7	975	6	AX652811	AX652811 Sequence
C 42	39	2.7	1182	8	AK065644	AK065644 Oryza sat
C 43	39	2.7	25150	1	AB070945	AB070945 Streptomy
C 44	39	2.7	107870	8	AC105262	AC105262 Oryza sat
C 45	39	2.7	133271	8	AC104277	AC104277 Oryza sat

# ALIGNMENTS

RESULT 1  
PROIRM/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

PROIRM 3889 bp DNA linear BCT 16-FEB-1994  
P.stuartii PstI restriction and modification genes, complete.  
K02081

K02081.1 GI:150922

endonuclease; methylase; modification enzyme; restriction

endonuclease.

Providencia stuartii

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Providencia.

1 (bases 1 to 3889)

Walder,R.Y., Walder,J.A. and Donelson,J.E.

The organization and complete nucleotide sequence of the PstI

```

restriction-modification system
J. Biol. Chem. 259 (12), 8015-8026 (1984)
84239756
6330092
COMMENT
Original source text: P.stuartii DNA, clone pPst201.
The two genes for the PstI restriction-modification system are
encoded on opposite strands and are transcribed from separate
promoters. The restriction endonuclease contains 326 amino acids.
The modification enzyme contains 507 amino acids. The close
proximity of the transcription initiation points of the two genes
(70 bp) suggests that the promoters overlap. The -10 regions for
(the two) genes are found at bp 2307-2302 (MME) and 2352-2357 (PstI).
Bases 2267-2340 are protected from DNaseI digestion. The PstI
restriction-modification system is a type II system. The two
enzymes both recognize the hexanucleotide sequence 'ctgcag'. The
modification enzyme protects this sequence from cleavage by PstI by
methylation of its adenosine residue. PstI cleaves the DNA between
the adenosine and guanine residues, leaving 'sticky ends' of 4 bp.

FEATURES
            source
            1..3889
               /organism="Providencia stuartii"
               /mol_type="genomic DNA"
               /db_xref="taxon:588"
               complement(<742..2292)
               /product="MME mRNA"
               complement(742..2265)
               /note="methylase modification enzyme (MME)"
               /codon_start=1
               /transl_table=11
               /protein_id="AAA25672.1"
               /db_xref="GI:455323"
               /translation="MTEAATQIPISLNIILVDSIREAANSTLDLTRSLKIQFQMSSSAV
               SELMANLPFSYGEHIEDLAGVGSILTAFAVQNATLNGAKSISSTCYEISEWVYNL
               IOVLDCIKRAMEFGEVNWOOKEIESDFIOASVEQLLIENYSPKYNAILNPYFKLIA
               KGRERALLQGVIEASNLVSFAVALAIKOLKSGSELVAITPRFCNGVPYDFRQKML
               DCSLNKHVFNRSKSAFADNVLQENIYHLTKGTORKVTVYVSTCANINDPTIF
               EYVFDSEIVSNPLDHIHVINEQRELANKAGLPCSLDGLGIQVSTKVDYDFRE
               NLSYMSVNSPLIFPQHLQRCISVWPIITAKKPNALIVNEATNNLWPNVGIYVLTFR
               LTAKEKRRIASIVYYPIDVAVDTGFDKNINYFHANGKPLDISLAKGLWVFNLSLI
               DKYFRMONGHTQVNAITDLRALRYPTREQLDIANQVDFGEFEQTKIDEIINOSLQM
               2363..>3376
               /product="PstI mRNA"
               2396..3376
               /note="PstI restriction endonuclease (PstI)"
               /codon_start=1
               /transl_table=11
               /protein_id="AAA25673.1"
               /db_xref="GI:150924"
               /translation="MKELKIKAEKILKALGPLPQOVNDRSGWVLLALANIKPEDSKW
               EKAKPLPVTSIMEFRTBYGKDYKENSRETIHQTLHOEQARIYDNRDLPSRATN
               SKDNYSNVNIIIDHYNPNWKNELIQOILHVSLELYERALARDRPIKLLDGG
               TQISLSPGHNQIHADIYHVFPRFVDMGKILYIGDTASRNEGKLMVLDSEYLRK
               LVPYPMSHDKLPVYVDEKRWKFLIEAVTSHGPIISPKRWLELEALSSCTVGKVVY
               TAPPTWTEFRKNAANIAWETVMIADNPDMHVHFGDRFLGPHDKKPELS"
               1 bp upstream of HindIII site.

ORIGIN      1 bp upstream of HindIII site.

Query Match      4.9%; Score 72.2; DB 1; Length 3889;
Best Local Similarity 45.8%; Pred.No. 9.2e-07;
Matches 481; Conservative 0; Mismatches 528; Indels 42; Gaps 5;

QY      400  ATTGCTAATCCTCCCTACGGAAGCTTGCCTACAGATTCCGTGGCGCGCTTGCAACGACA 459
      |||
Db      1818  ATTTTAAACCTCCATACCTTAAATAGCGCTAAAGTCTGAACGAGCCTTACTACAA 1759
      |||

QY      460  GCGCGTCCGCTGCAGATGATACCGAAGCTTTACGTGGCCTTCGTGGTGCAGCAGTCATTTCG 519
      |||

Db      1758  AAAGTTGGTATAGAAGCAAGTAATCTTTATTCTGCTCTTTGTGCTCTTGTTAATAACAG 1699
      |||

QY      520  CTCAAAGACGACGGCGGGGGGTTTTCATTGCTTCCTCGATCTTGGGCGAAGCGGCCCTTAC 579
      |||

Db      1698  CTTAAATCAGCGCGTGTAGTAGTGCTATTACTCTCGTTCATCTCTGTAATGGAACCATAT 1639
      |||

QY      580  TATCGTCAATTTGCGCAATTTGGCTGATGACCGCGGTAAAGTCTCGATATACTATGCTGTC 639
      |||

```

Kishida, Y., Kohara, M., Matsumoto, M., Matsumoto, A., Muraki, A., Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M. and Tabata, S.

Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120

21595285

11759840

2 (bases 1 to 343250)

Kaneko, T.

Direct Submission

Institute (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan

(E-mail: kaneko@kazusa.or.jp)

URL: <http://www.kazusa.or.jp/cyanobase/>

Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

Location/Qualifiers

1..343250

/organism="Nostoc sp. PCC 7120"

/mol\_type="genomic DNA"

/db\_xref="taxon:103690"

/note="synonym: *Anabaena* sp. PCC7120"

complement(188..6253)

/gene="all3691"

complement(188..6253)

/gene="all3691"

/note="ORF\_ID:all3691"

/codon\_start=1

/transl\_table=11

/product="serine/threonine kinase with two-component sensor domain"

/protein\_id="BAB75390.1"

/db\_xref="GI:17132825"

/translation="MESVMNTYVDLAIAPGYFILEIETIYHSGKTVVRAVREAOQOP  
WIKLLREYFTSLLQFRNQYAIKHLKIPGIVNLYSLPEYRNSYAMVEDFGDIS  
LRDYAQOQSLSTELITITQADILHHYQORIHKDKIPANILINPETKOVKLIF  
STASLLPRTONIISPNLEGTAVLSPQETGRMKRGIDYSDFSVLSGLVPELLTGE  
LPFTTDDMELVHCHIKQPKFTIPSCQRYANRFGSGSTKLPEETIPQVGE  
IYMKLMKNAEDRYQNALGKLYDKLCLAELOETKGYFTIGSRDVCDFRIFIPKLY  
GREAEITLAAFRNVSIGATELMSVAGSGIGKTAIVNEVHKPIVRQGYEIQKQFD  
QQRNIPFSAFYARNLQGLTSDTQIQMRSKILEAVGKQGVIIIEVPELEI  
IGKPAAPKLSGTSIQNRPNLLPKFTQVFTSAEHPVLVFLDDQLQWADSGSLKILQLL  
MNDTGVLILGAVRNEVNPAPLMTVNEIGKNGTIIKIDLA PLNOLKNTLVLET  
LKCSSETAQSALLVQKQNGPEFTTQFLKALHODNLIQFNYDLGQWCDIAOVNHQ  
ALTDVLFMAQSLLKRLRATQEVQLAACIGNQDFLELAIWESSPMKTAACLWA  
LQEGITIPQSEYKPFVQGEIHDHQAQTEIIVTYLHQRVQQAAYCLIPAEALIAH  
HHIGQLLQKISPAAREEHIFEIVNQLNYGITLIQQSERDELAQLNNAQKARAT  
AYQASREYADTSMTLGETAWQRYQLTITHELAELIALLCGDEVQNNQLTETVNH  
AKCPLDROYQVKIQAITSRNELLEAIATGKSLQELGSLPDHPTPEDVQQAAREEI  
NRLIGDRLGEPINLPKMTDPEKLAIMOISSLIIPACYMTGSLVLLIIVPQVKLSLQ  
FGNSLPSAGYVSAYPHLSTTQWMAQAQQLQAIAYQALASEPAKNIRAAATFVLGGY  
FYHCTSHLDITLPIIQEGCIAGETGNLEFFGYNVQNGNAYWSGESSEFASQMT  
YHQQLDNLNLTNSHCFYVWESALILIGKSADEISLRQDAYAEKIVTEALASNDLR  
LFQFYLRVLNLLADPRADQAEKARQYLTCMGSLVAEPIFYFYDSLTALAEHQ  
PAGRDSQWRVRENAILDKWADHAPMNVLHKWGLVAAETHRLLOKTEAKMTNLEKR  
YKAKENLRLDERALAEAAKLYLDWKEKVAAYVMQYAYCYAREAKAKTNDLEKR  
YPLLQKLIQERQKFEQNETIATVPGTSSLTSSLTGTSISDTDLFASILKAAQVIS  
TSLEDELITNTEILLENAGKSKALLPQEDIMQIKAMTILSNFQPSSESSTQILE  
SQPLECEBIPKNIIVYKNTQQTVIDNLQTDIFGLIYMLQHQPOQSVFTPMNQ  
GHVLGILLENLRTGVFTSDRLVILISQAASVLSLEKARYQESQTKAQQLKQSLK  
QOKILFNVNQMRQSLDLMALICVVTQNIIRILTDVDRVGIYQFHLDDVNYEYGFVAED  
VSPAPSAVAVQDHCNGENYANLYKQGIICAITDVOSSEILDKHROIQAQHVRA  
LVVPIQBEELMGLLCHICDPRPWEPLQFAQVQAGMQIALKQTDLLIQTQQA  
TOLEHTLQHQLOLQVONERKMSAGNLVAGVAHEMNNPLGFIASVLSQHTQPTFADI  
MEHLKLYQSLPHPGDKILHAAQIDLDYILEDPLKIDAMVMACDLKNIISTLRTF  
SHADYDKIPINFDIGDITLILKHLKANDKRAPIEIVTEYGLNPHIECPGQOLNQ  
VPMNLANADAEDANGLSYAEITANNRIIIRITOVNVHRLISITDNGIGMSEEL  
KQHIPEHLFTTKGVEKGTGLGLAIARQIWEKHGGAIEVNSLQSQGTEFVILPITAE  
SANKY"

6804..7658

/gene="all3692"

6804..7658

gene

CDS

/gene="all3692"  
/note="ORF\_ID:all3692"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75391.1"  
/db\_xref="GI:17132826"  
/translation="MNFQIQGVALLLTQGLKVLGAIALMWIAQKLIETAFKLIVRR  
GLRTONIEPTLISYLLNIVSVTLRIVLVAILGFFGIETTSFAALLAAGVAIGAAMG  
GILANFAAGAFILVPRPEKVGDFITAGITGTVSEIGLFTTITTPDNLVTIVANKI  
FADNIONFSANPFRFRVDLQAQHDVDNQAIALKARISQIPNVIENPAPDVEILTF  
NLAGEVLAVRYPYCNNDHYQVYFDTNKAIREFGEAGYPIPEQRYGFSVPENGTSSV  
IPSSALSS"  
8079..8939  
/gene="all3693"  
8079..8939  
/gene="all3693"  
/note="ORF\_ID:all3693"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75392.1"  
/db\_xref="GI:17132827"  
/translation="MNVAEISQVALALITQFGLRVLGAIALMWIAQKLIETAFKLIVRR  
AFRSKHVEPTLITNYIVNIVDVLKIVLIVAILGFFGIETTSFAALLAAGVAIGAAMG  
GILANFAAGAFILVPRPEKVGDFITAGITGTVSEIGLFTTITTPDNLVTIVANKI  
FADNIONFSANPFRFRVDLQAQHDVDNQAIALKARISQIPNVIENPAPDVEILTF  
NLAGEVLAVRYPYCNNDHYQVYFDTNKAIREFGEAGYPIPEQRYGFSVPENGTSSV  
VQRTSLGRG"  
9664..9831  
/gene="all3694"  
9664..9831  
/gene="all3694"  
/note="ORF\_ID:all3694"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75393.1"  
/db\_xref="GI:17132828"  
/translation="MAVDLDKRLVLILQSFHLRSWERNGLWEVSRRIKETHVMS  
VAPEIELVIPG"  
complement(9841..11007)  
/gene="all3695"  
complement(9841..11007)  
/gene="all3695"  
/note="ORF\_ID:all3695"  
/codon\_start=1  
/transl\_table=11  
/product="N-acyl-D-glucosamine 2-epimerase"  
/protein\_id="BAB75394.1"  
/db\_xref="GI:17132829"  
/translation="MGRNLQALQALYKVALINDVLFPWENHSLDTEGGYFTCLDRQK  
VDEDFKFWLQNRQVTFMLCNQLEKRENMLKIASNGAKFLAQHGRDAEGNWFALT  
RGDELPQFYIFSDCFAMASQYALASGEAWKAVAMAYNNVRKDNPKGYTK  
TYPRGRKALAPMILANTLEMEWLLPQETLENVLATVOEVMSDFLDKRGMYE  
NVAPDGHIDCFEGRILNPGHICEMWFTMDIARKNDSKTINQAVDVVNLTNFAMD  
KEYGCLYFEMAAGHPROOLEWQKLMWHLSELVALAMGYELTGREYCWQWQKMD  
YSWQHAPDEYGEWFGYLNRRGEVLNLLKGGKWKCGFHVPRALYLCWQQFEALS"  
complement(11134..12273)  
/gene="all3696"  
complement(11134..12273)  
/gene="all3696"  
/note="ORF\_ID:all3696"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75395.1"  
/db\_xref="GI:17132830"  
/translation="MLPQALDNGRLLVAENGLTQOVLIVYKIQDQPVQVQTFQHTG  
GIYGVPGVQDLKFGYGTGADAGQGLNLYNNNGFNNSGTDLRKFSQKQWLKLLG  
LSLVENADADPKTDGREFVTKYQFLIDSSKPKQQLIYKAYITLNGKYPQDSRLHIS  
PDTTFVRINDERFLFLDDPSNALEIYRPNPKDGNIAIPAGMEVGNIRGKSAISG  
TWPPQEPGTGKIWNDRNGDGAFDKEEYDRGEDSSSVTGMWVDSKGDVWTTILGDRRG

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene  
CDS

RHYPLQGLDNGNPIYVNSIQEKTPLIFTLRRRIKYFPESDTMVLGFTAKNPPFA  
EDQAPGSEIACFDNWSKDNILRWRIVVPKDIRKEMITASTH"  
complement(12278..13114)  
/gene="all13697"  
complement(12278..13114)  
/gene="all13697"  
/note="ORF ID:all13697  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75396.1"  
/db\_xref="GI:17132831"  
translation="MKHEILDEFKRLQSRRAKOSNKIFKVLVSLRLKSWIFLIIAIS  
FLATTITADTLSEVAENSDKDLMDIFADSKQSIQNYIEGIVYVAPNGVTVTNSHE  
DKAGASIVKDGNGVIGVSLDGMSPRGKAVTANSKYLVIAMSQGFVKIDKDP  
EGTWYCVRYNLSGRPAPFANGRWKSMILVNTKSEVIGLATKDKRLVSDAANK  
IRVYNTETMELRNFSFLRPEDITIDKQNLWIIQKODANNPAKILRYSSGKQLPOK  
NN"  
14028..15197  
/gene="hepB"  
14028..15197  
/gene="hepB"  
/note="ORF ID:alr3698  
hepB gene product"  
/codon\_start=1  
/transl\_table=11  
/product="heterocyst envelope polysaccharide synthesis  
protein"  
/protein\_id="BAB75397.1"  
/db\_xref="GI:17132832"

gene

CDS

translation="MEYKREKLIVSASILTLGSMPTNPGLIERYIYLTQLSAN  
QDVELCGVLGNOPNOLPIKLTNLASPSDKIWOFWSIRNFKOTRIGKPDANLHF  
ALYSFPLDILPOGIPITFNHFGPWAESKQELVKNKISIFELKRLIETQVYHNCDF  
IVLSKAFNHLQYQIIPWKIHIIPGVNIDKPNLSRQARQALNWPESRILFT  
SRRLHVRGVNKLQALAIKPKLIDWLAIGRHQLTLEKQALQELGNVNVKFG  
FLPDEQLPIAYAAANTWMPQSFEGLAITESLACGTPLCTPIGMEPIILTPSP  
QLITASPEATAIAEKIAILLQIETPKPSRECRQYAVTNFDWKIAQQVRVILA"  
15259..16407  
/gene="alr3699"  
15259..16407  
/gene="alr3699"  
/note="ORF ID:alr3699  
similar to glycosyltransferase"

Query Match

Best Local Similarity 4.18; Score 59.4; DB 1; Length 343250;

Matches 270; Conservative 0; Mismatches 281; Indels 12; Gaps 3;

Qy

865 TTTACTCTGCTGATCGGCATCGGTGAAGTACGGGAAAGGTGTGTGATTTTCGCAAT 924

Db

17439 TCTACCTTAGATGAACCTGGTTAGCAGTGTCAACGGGCCAGTTGTAGATTTTCGCCTC 17498

Qy

925 CGTCAGTATTGACCGATAACCTGGATGCTTCAGCGGTTGTGCCCATGTTTATCAGTCA 984

Db

17499 AAATCAGCTTTGAGAAATATGTAATG---AGGAAACTGTTCATTTACTTTACCCAGAA 17555

Qy

985 AACATTCGATCTGTGTAATTTGATTTGGCTTCAGTGGTGGAGGAAAGCTCAAGGATTT 1044

Db

17556 GCATATAAACAAGGAAAGTTTATTTCCGCCAAAAGCGCGTAACTCATCGCCATT 17615

Qy

1045 GTTTCGGTTCGAAGATGTAGCACTACGTCAACTTCTCCCGCAAGGCGTGTATGTTGTG 1104

Db

17616 GTACAAATCAAGAAACAGACAAATGGCTAAT---CCATCAGGTGTGTATGTTTTCACG 17672

Qy

1105 AAACGCAACGCGGAGGAGGACCGTCTGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1158

Db

17673 AAGCGATTTTCTGTAAGAAGAAACAAACGCGGGTGTGTGCTGTGTATGTTCTTCCTGTA 17732

Qy

1159 GGGCGCAGCAGGGTGTGGCTCGCAATAAACAAGCAACTATTGTCATGAATCTCAACGACCG 1218

Db

17733 GATGCACCGAGTGTGGCATAGAAATCACTCACTACTACCACTTCTCAAGGTGAGGGA 17792

Qy

1219 CTTGAGAAAAATGTGGCCCGGGCCTCATGCTTTGGTTGAATCGACTGCTGTGGATCAG 1278

Db 17793 ATGAATCCCGACCTGGCAACGAGGTTTAGCAGCAATTTCTCAGTTCAACTTTATTGGATAGT 17852  
Qy 1279 TATTTCCGAGCCTTTTCCGGGCATACCCAGGTGAACCTGGCGATCTACGCCGCTTCCG 1338  
Db 17853 TACTTCCGCTATTTCAGTGGACATACACAAGTAATCAACGAGTTTTCGCTAGAAATTAAG 17912  
Qy 1339 TTCTCTGTGCGGAGGACCTAAATTTCTTCGTAAGGTCTCCCGATGGCCTCGCTGAT 1398  
Db 17913 TATCTTGAAGATGATTTAATCAAGCTAGTAGTCAAAATGGTGAATCTTCTGCTGAT 17972  
Qy 1399 CAGGAGCAGTTGGATGCCGTGT 1421  
Db 17973 CAGGCTCACTTGATACGGTTGT 17995

RESULT 3

AF146609

LOCUS

DEFINITION

partial cds.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

FEATURES

source

CDS

1..1600

/organism="Aeromonas hydrophila"

/mol\_type="genomic DNA"

/strain="PDI34/91"

/db\_xref="taxon:644"

185..>1600

/codon\_start=1

/transl\_table=11

/product="putative modification methyltransferase"

/protein\_id="AAF45040.2"

/db\_xref="GI:7327641"

/translation="MSSLPENIKHVNLLDPGGPGSLTAFAVDEAIRGLTSLSIH  
AHDIEERTIPFDIKTLQACNAGIDATHKFLNDFLINKTSDSKNYSVKYTHC  
IMPPYKKTASSSEHRRKRLHAGIETVNLVAAVSLVONLSGGELVAIIPRSFCNG  
PYQPPRELILRETSIKHIIFESRNAPSDDVLOENIIHLIKNKQSDVITTSPP  
IADFLHDDTGSVTASDMTIRTVPFSSKIVSAQKQFIIHANNRQELIDRLNSFA  
TLEDLNIQVSTGVPVDFRLKEDLRQNIPEGAVELIYVHLNNGNSMPKSKKPNAMV  
SEGSKWLKWHGHGFIVRRFSKKEKRIVATNYDSCLPGLIGENIKFNKLNPHANKS  
GIDRLAFGLTIYLNSTLDKYRFLGGHTQVNNATDLRNHIHYPSSINSLRMGQVDNL  
DLTKQDIDGLDKELDMTGEN"

ORIGIN

Query Match 3.58; Score 51.2; DB 1; Length 1600;

Best Local Similarity 55.7%; Pred. No. 0.13;

Matches	98;	Conservative	0;	Mismatches	78;	Indels	0;	Gaps	0;
Qy	1172	TTGGCTCGACAATAAAACGAACACTATTTCATGATGAATCTCAACGACCGCTTCAGAAAAATG	1231						
Db	1317	TTGGTTTGAATAAAGCTTAATGTATTTCATGCCAATAAGATGGAATAGACCGAGATT	1376						
Qy	1232	TGGCCCGCGGCTCATGCTTTGGTTGTAACACTCGACTGTGTGGATCAGTATTTTCGAGCCT	1291						
Db	1377	TGGCCTTTGGTTTAAACGATTATTTTGAATTTCACACTTTAGATAAAATATTACAGACTTT	1436						
Qy	1292	TTTCGGGCATACCCAGGTGAACGCTGGCGATCTACCGCGCTTCCGTTCTCTCTGT	1347						
Db	1437	TTGGTGTCTACACAAGTAACGCTACAGACCTCAGAAATATTCAATACCCAAGT	1492						
RESULT 4									
AX655393									
LOCUS	AX655393			2000 bp	DNA	linear	PAT 22-MAR-2003		
DEFINITION	Sequence 5263 from Patent WO03000898.								
ACCESSION	AX655393								
VERSION	AX655393.1			GI:29158207					
KEYWORDS									
SOURCE									
ORGANISM									
	Oryza sativa								
	Oryza sativa								
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;								
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;								
	Ehrhartoideae; Oryzeae; Oryza.								
REFERENCE	1								
AUTHORS	Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,								
	Katagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.								
TITLE	Plant genes involved in defense against pathogens								
JOURNAL	Patent: WO 0300089-A 5263 03-JAN-2003;								
	Syngenta Participations AG (CH)								
FEATURES									
source	1. .2000								
	/organism="Oryza sativa"								
	/mol_type="unassigned DNA"								
	/db_xref="taxon:4530"								
ORIGIN									
Query Match	3.4%;			Score 49.6;		DB 6;		Length 2000;	
Best Local Similarity	8.4%;			Pred. No. 0.33;					
Matches	75;			Conservative 395;		Mismatches 414;		Indels 4;	
Gaps	1;								
Qy	144	CCGGGAACCGTGGGTGCTCTGACACCAGAGCTGTGTGGCTCTCTGACCGCTGCCCT	203						
Db	29	CARMSSRSRKMGSMSKYKCSGCKKWTTRKSKWYSASSASRTSKWSSGYSYSGK	88						
Qy	204	CGTCGATCGCTGCATCTAGAAACCCCGCAGCTGGTGGCTCAGTGTGCCGTGGAAC	263						
Db	89	GMKKRYKRSGRGRGRGRMRMRMRMRGRYRCARSGRMAGSGRMWGKSRMSYMMC	148						
Qy	264	CGACCCCTTTGTCGTCTTACCTGCGCGCCACCTGGAGGATGTCGNAACGCTTACGG	323						
Db	149	YARGCSGRKKSGSGWGKTCRRGARGSGWSSGAKYKSGSKRMWMSOGRSGCGR	208						
Qy	324	CAPCTCCTACGACTAGTCGAGCGCACTATTTCCTTAACAGGGGCCAAGCTGGATCG	383						
Db	209	SAYSRYGTSRKYGTYYKMTTYSASRCMAYMTTYSWACSSYTWCRSRKRRMMWOMRK	268						
Qy	384	CCGCTTCGATCTTGTAATGTCTAATCCTCCCTACGAAAGCTTGCTCAGATTGCTGGC	443						
Db	269	MRWSRYGYSWYSQWMMCTAYKYSYRWCYMYRGGWGRGATRYWGRGYMSRMAMMYK	328						
Qy	444	CGCGCTTGCAACACAGACGCGGTGCGTCGATGTACCGAAGTTTACGTGGCTCTGGGT	503						
Db	329	MYWRYGKMGKRWAGRMMSRCWRSKACAYMYRWRMRMTTTRRRWAKSRSTSRKKR	388						
Qy	504	CGCAGCAGTCATTTCGCTCAACAGACAGCGGGGGGTTTTTCATGTTCTTCGATCTG	563						
Db	389	KWCMRKRYKRMGYSRMSCKRWRWKRSGRAWMGRCGMTCRMKYSGMWRKWSK	448						
Qy	564	GCGAAGCGGCGCTTACTATGTCGTAATTCGCCATTGGCTGATGACCGCGGTGAAGTCTG	623						

[illegible]

```
/translation="WTOILETVDKSRILTVPNLLKNSKSELQGFQFTTSSISIFMACLFSE
DKLNNKVLADAGAGISLTSFAFLARLISENIGKADLHLIDEMLEPYLGETLALFKD
YIELNSQILIDDFENWAXSLIDDESLANDKOPFTHAILNPPVKLIKSKHEKRLAR
KAGLETNLYSAFVALVDMSOGGEIVFIIPRSFCNGPFRHFRHLLNKTSLKHHM
LFESRDKAFKDVNDQVLSKUEKGTQVEDVKISISTDDSFVSIRYRPFKEIVOP
NDIEKFTHINTNEETLIEKPNVICYLEELNIEVSTGPVDFRKNELREMPGEIV
PLFYPNHFVGTSLSEYPMKMKPNALIRNEKVKRLYPNGHYVVVVRFSSEKEKRIIVA
GLTPESVNDPVGVFENGLNVHYNKSGISKEVAYGLVAYLNSVFDVKYFRIFNGHTQ
VNAIDLRMFPSRDILISLGKGVNIENIVNGQVEIDSKLEELLNDRGNA"
1697. .1701
/gene="hsmBI"
1714. .2664
/gene="hsmBI"
1714. .2664
/gene="hsmBI"
/codon_start=1
/transl_table=11
/product="restriction endonuclease"
/protein_id="AAA18170.1"
/db_xref="GI:143055"
/translation="WTRGMSHNVKEATIKELGLPKGOONERSALCLLSMNIITODK
TWSAESPLGITPMFEFRINRYKEVAPNSRETFRRPTMHQFVDAGIALVNPDKPFR
PVNSKAVYQIEATLIEKYNTEESSELLARLYLSNRQTLVRYAKERQONKIPVQI
AEGKEIYTPGEHELKATIEBFAPRYVPGRLTYAGDTGKMGYFDEELRLGVV
IDSHGKMPDVVIVYPEKKWLLILIESVTSHGVDHKEELAKLFNGSTAGIVYVTAFF
NRSLLMARYLNNISWETEVWADAPSHLIHFNGVRFLGPE"
ORIGIN
Query Match 3.1%; Score 45; DB 1; Length 2892;
Best Local Similarity 49.5%; Pred. No. 4.7;
Matches 154; Conservative 0; Mismatches 145; Indels 12; Gaps 1;
Qy 1075 CTTCTCCCGCAGGTCGTATGTTCTGTGAAACGGCAAGCGGCAAGAGACCGTGGT 1134
Db 1332 CTTTACCCTAATGGGCATATGTTGTTTAAAGGTTTCTTCAAAGAGAAAACGT 1391
Qy 1135 CGTGTCACTCGTCGGTCT-----GGGACGGGGCGGACGAGGTTGCGGTCGAC 1182
Db 1392 CGAATAGTAGCAGGAGTATTAACACGAGAGGTGTAACGATCCAGTAGTTGTTTGA 1451
Qy 1183 AATAAAGCAACTATTTGCATGAATCTCAACGACCGCTTGAGAAAATGTGGCCCGGCGC 1242
Db 1452 AACGACACTCAATGTACTTCACTATATAAGAGTGAATTTCTAAAGAAAGTAGCGTATG 1511
Qy 1243 CTCATGCTTGTGTTCACTCGACTGTGTTGGATCAGTATTTCCGAGCGCTTTCCGGGAT 1302
Db 1512 CTTTATGCTATCTTAATCTGACCCAGCCAGTAGATAGTATTCAGAAATTTTAATGTCAT 1571
Qy 1303 ACCGAGTGAACGCTGGCGATCTACCGCGCTTCGTTCCCTCTGTCGGAGGACCTAATT 1362
Db 1572 ACTCAGGTAATGCTACTGACCTTAGAACAAATGAATTTCTTAGTAGAGATATCTTAATA 1631
Qy 1363 CTTTCTCGCTAA 1373
Db 1632 ACTTAGGTAA 1642
```

```
RESULT 6
U67554/c 10929 bp DNA linear BCT 28-JAN-1998
LOCUS Methanococcus jannaschii section 96 of 150 of the complete genome.
DEFINITION U67554 L77117
ACCESSION U67554.1 GI:2826365
VERSION
KEYWORDS
SOURCE Methanocaldococcus jannaschii
ORGANISM Methanocaldococcus jannaschii
Archaea; Euryarchaeota; Methanococci; Methanococcales;
Methanocaldococcaceae; Methanocaldococcus.
1 (bases 1 to 10929)
REFERENCE Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
AUTHORS Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A., Tomb,J.F.,
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,
Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
```

```
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
gene
CDS
gene
CDS
gene
CDS
```

```
Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M.,
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
Smith,H.O., Woese,C.R. and Venter,J.C.
Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii
Science 273 (5278), 1058-1073 (1996)
96337999
8688087
2 (bases 1 to 10929)
Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.F.,
Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weidman,J.F.,
Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M.,
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
Roberts,K.M., Kaine,B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
Smith,H.O., Woese,C.R. and Venter,J.C.
Direct Submission
Submitted (27-AUG-1996) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jan 30, 1998 this sequence version replaced gi:1591751.
Location/Qualifiers
1. .10929
/organism="Methanocaldococcus jannaschii"
/mol_type="genomic DNA"
/db_xref="taxon:2190"
180..710
/gene="MJ1109"
180..710
/gene="MJ1109"
/note="similar to SP:P09556 percent identity: 38.37;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="orotate phosphoribosyl transferase (pyrE)"
/protein_id="AAB99112.1"
/db_xref="GI:1592320"
/translation="WDKSKLNLKLVGTCIRFGFEFIFLASGKSKSNYYIDIKATTNPE
ILKVGELIAEQIDKEDVKAGVELGVSPIATAVSIIAQPLLIVRKPKDVGTKNKI
EGLKRGKGVIVEDVTGGSVLKAVKEIRENGGIVDKVFVVVDLEGAKENLQKEN
VELPLVTVKELQSTQ"
complement (808..2268)
/gene="MJ1110"
complement (808..2268)
/note="similar to GB:AB000666 percent identity: 30.00;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB99113.1"
/db_xref="GI:1499961"
/translation="WESTSLIKVANEFLDSINEINDKLKEVVGKIKNKTIKTKLSD
IILLEKLEILQDLKSKMEFLFDSYKVVNKGKGGVDSGLQELASTSYLRRIAS
BKGLILERVHARLVAHIALHTEDIGNINLPNPLDGSYKIMFEPPVLYVTYK
RFLDLEPKRGILTSYISLVIDKGRFKRIVKEDQYKYEIKFEKGNAIITSIK
RNFSKNLIIDQYVVRVLAIGVNTYKDEIKANEIKDKLNEEKYLNRYLELCL
LFREADISGGLDVRCEERKLELEKLEKGLYRDGPEIPLKKAIKKNELS
KKISKDIIKRPSEDVFRFYLYKTTPDERASNLFPFSIMTTPQGFLSWMSVDGINVD
VLDFKLEELPKYQIFLNKIGVALYLHWDADVERFNFKKKDIEDLLKXIALIEP
IKELDKNDVDVSKLEKFGKVKETKFLDLLSL"
complement (2231..3580)
/gene="MJ1111"
complement (2231..3580)
/note="similar to GP:1575544 percent identity: 38.14;
identified by sequence similarity; putative"
/transl_table=11
/product="tRNA nucleotidyltransferase (cca)"
```

```

/gene
CDS
/protein_id="AAB99114.1"
/db_xref="GI:2826366"
/translat="MIVLTIEILKEVLNTEIKPSDEMEKILQKANEILIDKIWEIVRE
NSYLVASHVGVNGEVGYEDVIVPCVILDFEKEILISAVDKTPUHHKFLNRUNERUC
DEVRLKAFKLSGSDVTKGFSGLCELLILHYGSFNILLKEAQNWRIGKRIIL
KDFEIVYKVDINDKLEKFEFFIYVDPVLDNRNVASPLSKDNFCRFIYRQFLKNPS
IEFFKAYKADILEYELVYELPKIALREGPPVFEKERAREFLAKYGVKVFIRDCCK
VLANSKCFADDDNYLWYELVYELPKIALREGPPVFEKERAREFLAKYGVKVFIRDCCK
LFATERYEYSHIDLEKDIVNGNLQNTISIPKYVNPNGKIELNSHGEHKEQKCEQ"
complement(3699..4604)
/gene="MJ1113"
complement(3699..4604)
/gene="MJ1113"
/notes="similar to GB:D28748 SP:P39465 PID:506372 percent
identity: 28.00; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="N-acetylglucosamine-1-phosphate transferase"
/protein_id="AAB99115.1"
/db_xref="GI:1591753"
/translat="MGHYFNLTYYTIIAFISAVLCKFLMKMKINKYFGYDLHKKEK
IKVPMGGLAVLSNLFIFVPMPIFVLPITTAGIIGVDDIAKLSPEKLIILFISG
LIIGILFNNSYNLIEILIALGIMISSLTNNLAFNGLEIGCMGVIASISLALVLF
LDNYTTFGLSALIFSASYLGLIFNKYPKVPFDPVGLTPIGAFVLAVLVVKEYIYIP
LVIMPPYVIDASLKYSAGVMSRDEKPTTLKEDGKIYIGGYLSLPLRLIKYKPMRE
PLVTLVIIGIPIFGVIGILSLIA"
complement(4910..5488)
/gene="MJ1114"
complement(4910..5488)
/gene="MJ1114"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ1114"
/protein_id="AAB99122.1"
/db_xref="GI:1499965"
/translat="MDIVEKVKYEGILKLENIPQIILNVLVAGLIWVFGILVFIPIA
DMLGNPLFGLTALKYPIISAIITIALIIVLRVTKPFGELMDGIADIIAVKLGRVN
EELKYYRGLRGLAYLVAIIAYLFLPVISGITPVLAGIVLIIILVMAVTVLINIG
HIFSEIEGIRIATEKLEKESVKNEENE"
complement(5572..6357)
/gene="MJ1115"
complement(5572..6357)
/gene="MJ1115"
/notes="similar to GP:1652059 percent identity: 41.80;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB99116.1"
/db_xref="GI:1591754"
/translat="MIPMFKKPLIGMVHLKPLPGSYHYNDNFDDIVFAIKEAKKLE
EAGFADVMENFDGDAFKEADKITIASMAVIAKAIKEEVSPLGINILNRDAIGAYS
IAYVVKADITRVNLSVAFDQGITGKAVELAKLLPSKIKVFDVHVKHVHF
IDPESLLDTRVGLADAVIISGRTKGEVDIEKLAKELVDVPIVSGSTNNYLR
ILNSYADGFIIGTWIKDKKANNEIDIDRAKKIVNLANKLMKC"
complement(6372..8006)
/gene="MJ1116"
complement(6372..8006)
/gene="MJ1116"
/notes="similar to GB:J05554 SP:P22106 PID:145393 GB:U00096
PID:1651277 percent identity: 33.78; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="asparagine synthetase (asnB)"
/protein_id="AAB99117.1"
/db_xref="GI:1591755"
/translat="MMDMCSISGIIIVKDNQISAKYSIDMMKILKHRGRDNGSGLLDDE
VIYFNDPVEDLEEMIGNLSLAHRLAIVGRYGVQPIPNEDETIWLVCNGEINYI
```

```

ELREYLKQNHFRDSDNEVLIHLHYEBEKELEEDGDYAFATYDKSKNVRLARDMFGV
KPLFYVDKDYFAPASERKALWHLLINIDGCRDLDELNSKILTKLQNSQIYYIDDN
RFELIEGFKLEINMKERSYEAKYEDLRALKNSVLRVGLDKVGLICSGVDSSL
IAKLASYCEVILYAVGTENSEDIYAEIRLAKDLNLRKKIISBEEYEYVFKVAKA
IEDVLMKIGVPIYVASEMANEDGKLVLSGQADELFGYARHERIYRERGEEL
KKELLDMKYNLYKVNLRDDHCTMANGVLRVLPFLDEVVIEIALSIDIEYKMSLSNR
PYAESNISLAKSEPIGLKNTNLNICKVRSRKILRDVASQYLDPIYAIRPKKAAQYG
SGGEKMIYKAKKYGFSGKINEFLDLMKRKIVSEF"
8143..8733
/gene="MJ1117"
8143..8733
/gene="MJ1117"
/notes="similar to GP:1296941 percent identity: 40.49;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB99118.1"
/db_xref="GI:1591756"
/translat="MDALIMAGGKTMGGVEKPELKLGRCLIDYVVSPLLKSKNN
IFTATSPNTEKTEYINSAYKDYKNIVVITDSKGIEDNECIGFSEPLVVSDDL
INJSLIINSVIDFYICIKAKTPDEALAVMIPKYPNPSIDFNGLVPAGINWSEK
HGQKBEIHWVIDLIFNINTKDDLKLAEMLLKKDGL"
8758..9027
/gene="MJ1118"
8758..9027
/gene="MJ1118"
/notes="similar to GB:AB000782 percent identity: 32.88;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB99119.1"
/db_xref="GI:1591757"
/translat="MEKMPAKLFFERSIIGKSGVIGKVDIVFDEKVGRLVSEVEP
AEHSPIMREGRNVLIPYLIKVAIKDWWIDETNLNRVIRVAEH"
9141..10028
/gene="MJ1119"
9141..10028
/gene="MJ1119"
Query Match 3.1%; Score 45; DB 1; Length 10929;
Best Local Similarity 47.4%; Pred. No. 5.3;
Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 681 AGAAGATGTCATCGTTCCTTTCAGTGTGAGCGCGCAAGCTCTAGTGTGCTTCTAG 740
Db 6019 AGTTAATGCTTATCTGCTGTTGCAITTCAGACCAAGGATTCAGAGGCAAGCTTA 5960
QY 741 GTCGGTCCACATGGAGAGAGTCGATCGCAAGTTCCTGCGGTAAGATAAGGTTTTCAGATGTTCA 800
Db 5959 TGAATTAGCAGCTAAAGAGTTGCTTCCAGTAAGATAAGGTTTTCAGATGTTCA 5900
QY 801 TGATGAAGACGATGACAAATTCGTGCACCTTCGCGGAAAGCGCATCGCTCGCCCTCGCGCGC 860
Db 5899 TGTAAAGCATGCATATCAITTTATAGACTTTGAAAGCTATCTTGGATACCGTTGAGAG 5840
QY 861 GAGTTTACTCTCGTGTATCGGCATCGGTGTAAGTACGGGAAAGGTTTGTGATTTTCG 920
Db 5839 AGGTTTGTGATGCTGTAAATTATCAGCGGTAAAGAAACGGGAAAGGAGGTTTCATTTGA 5780
QY 921 CAATCGTCAGTATTACCGGATAACCTGGATGCTTCAGCGGTGT 965
Db 5779 AAAGCTAAATAGCTAAGGAATTTGTTGATGTTCCAGTTATTTGT 5735
RESULT 7
AR271569_04
WPCOMMENT
Sequence split into 17 fragments LOCUS AR271569 Accession AR271569
Fragment Name Begin End
AR271569_00 1 110000
AR271569_01 100001 210000
```

AR271569_02	200001	310000	
AR271569_03	300001	410000	
AR271569_04	400001	510000	
AR271569_05	500001	610000	
AR271569_06	600001	710000	
AR271569_07	700001	810000	
AR271569_08	800001	910000	
AR271569_09	900001	1010000	
AR271569_10	1000001	1110000	
AR271569_11	1100001	1210000	
AR271569_12	1200001	1310000	
AR271569_13	1300001	1410000	
AR271569_14	1400001	1510000	
AR271569_15	1500001	1610000	
AR271569_16	1600001	1664976	
Continuation (5 of 17) of AR271569 from base 400001 (AR271569 Sequence 1 from patent US			
Query Match 3.1%; Score 45; DB 6; Length 110000;			
Best Local Similarity 47.4%; Pred. No. 6.5;			
Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;			
QY 691	AGAGAAATCATCGTTGCTTTCAGTGTGAGCGCGCAAGCTCTAGTGTGGTCTTCTTAG 740		
DB 16035	AGTTAATGCTTATCTGCTGTTGCTATTACAGACCAAGGATTTATCGAAGCGCAAGCTTA 16094		
QY 741	GTCGGTCCCATGAGAGAGTGCATCGCAAGTCTGCGGTTTCTGCGCTTCTGCTTCTCA 800		
DB 16095	TGAANTAGCCCAAGCTTAAAGAGTTCCTTCCAAAGTAAGTAAGGTTTTCGAGATTTCA 16154		
QY 801	TGATGAAGACGATGACAAATCGTCACTTCGCGGAAAGCGCATCGGTGCCCTCGCGCGC 860		
DB 16155	TGTAAGCATGCATATCATTTATAGACTTTGAAAGCTCATTTGTGATACCGTTGAGAG 16214		
QY 861	GAGGTTTACTCTCGTGTATCTCGGCATCGTGTTAGTACGGAAGGTTGTGATTTTCG 920		
DB 16215	AGGTTTAGTCTGATGCTGTAATTATCAGCGGTAAGAGACGGGAAAGGAGGTTGATATTC 16274		
QY 921	CAATCGTCAGTATTTGACCGATAACCTGCATCTTCAGGCGTTGT 965		
DB 16275	AAAGCTAAATTAGCTAAGGAATTTGGTGTGATTTCCAGTATTTGT 16319		
RESULT 8			
LOCUS YEN414030	5039 bp	DNA	linear BCT 27-MAR-2003
DEFINITION Yersinia enterocolitica Yeri Gene and ORF1 and insertion sequence			
IS1222 ORF1 (partial) and ORF2.			
ACCESSION AJ414030			
VERSION AJ414030.1	GI:17977862		
KEYWORDS methyltransferase-endonuclease; ORF1; ORF2; ORF3; transposase; Yeri			
SOURCE Yersinia enterocolitica			
ORGANISM Yersinia enterocolitica			
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
Enterobacteriaceae; Yersinia.			
REFERENCE 1			
AUTHORS Antonenko, V., Pawlow, V., Heesenmann, J. and Rakin, A.			
TITLE Characterization of a novel unique restriction-modification system			
JOURNAL FEMS Microbiol. Lett. 219 (2), 249-252 (2003)			
MEDLINE 22508023			
PUBMED 12620628			
REFERENCE 2 (bases 1 to 5039)			
AUTHORS Rakin, A.V.			
TITLE Direct Submission			
JOURNAL Submitted (21-SEP-2001) Rakin A.V., BAF, Max von			
Pettenkofer-Institute, LMU, Pettenkofer Str. 9a, Munich, 80336,			
GERMANY			
LOCATION/Qualifiers			
1. .5039			
/organism="Yersinia enterocolitica"			
/mol_type="genomic DNA"			
/strain="8081"			
FEATURES			
source			
db_xref="taxon:630"			
country="USA"			
1. .432			
transposon="IS1222"			
complement(<1. .71)			
/function="essential for IS1222 transposition"			
/note="ORF1"			
/codon_start=1			
/transl_table=11			
/product="transposase"			
/protein_id="CAC95148.1"			
/db_xref="GI:17977863"			
/db_xref="SPTREMBL:Q8VLS1"			
/translation="MLMCDATGLSQRRACRLTGLSLST"			
complement(98. .361)			
/function="essential for IS1222 transposition"			
/note="ORF2"			
/codon_start=1			
/transl_table=11			
/product="transposase"			
/protein_id="CAC95149.1"			
/db_xref="GI:17977864"			
/db_xref="SPTREMBL:Q46611"			
/translation="MKKRFSDQIISILREAEAGVSARELCRKHAISDATFTYWKY			
GGMVEPEVKRLKSLSEENARUKLLAEAMLDKEALQVALGRKY"			
389. .432			
/rpt_type="INVERTED"			
/rpt_type="TERMINAL"			
1100. .3580			
/gene="Yeri"			
1100. .3580			
/gene="Yeri"			
/function="methylation and restriction of DNA"			
/codon_start=1			
/transl_table=11			
/product="methyltransferase-endonuclease"			
/protein_id="CAC95150.1"			
/db_xref="GI:17977865"			
/db_xref="GOA:Q8VLS0"			
/db_xref="SPTREMBL:Q8VLS0"			
/translation="MLEVDEIRVKANALDNVKKGELGQFTSSSICIFMASLNE			
KGDISLDPGCGSLTAATFEETVRGARSLELHAIDIERKIPFLDVLVDKCSA			
SNAAGIKKIYQALINDYITAAQSVTHQDTEMYTHCIINPPYKKTISADYKILSAI			
GIEAVNLVAGFVALIMQLKQGEVMAIIPRSCNGPYLPFRNFIFQCAIKHVHIF			
IPRSHAFSEDDVLOENIIHLVNGIQESVITSPNSDFFDQDSNSVSDMTVRN			
IPRSLVNMLDKDFIHTAANNRQSIIDRLNVFYTSLNLCISVSTGEVDFRLKSD			
LRNIEPCAVPLIYPVHLNGVVDWPKSKKNAINVSRSRSLWNSGCGYFVIVRFS			
SKEKRRIVAIVIDGSLSGEMIGFENKLVNPHINKSGMDKDIAYGLSALNSMLDKY			
YELFGHTQINATDLRSLHYPDRKSLQIGSTISSQGLSQENINAINTEIKRLSKND			
DKNPAAQEKLDQALEIITLLGMPKSOQNSALTFLLALVNLNRPESGWOELKPLVG			
TPIMDCRDIYGEYAPNTRTFRTQLHRTDGLVLYNPKPRAVNSPKACYQIA			
PELFDVNTYGTPLMKNALGELWMORETLVEQYAMKRMHMIPLTIDNGTEIHLSPGD			
HSQLHDIYTFEGPFAFGSOVIXLGDGTAKEDPFRKDALADLGVTVNRKGLPDVL			
YWPQDWLILIESVTSHPGVGKRHSELANLEFKDARPGLVVYVSAPFDKTKMSKFS			
SWETEIVIAEAPTHMLHNGRFLGFHN"			
3790. .4251			
/function="unknown"			
/note="ORF1"			
/codon_start=1			
/transl_table=11			
/product="hypothetical protein"			
/protein_id="CAC95151.1"			
/db_xref="GI:17977866"			
/db_xref="SPTREMBL:Q8VLR9"			
/translation="WHIVQVSTSLILLITTPHTCINTLIDFPKNIQIKIMIEL			
RVCMTWEYKFXFDLHQALHSEVITSVKNAGIEINKKITDNGVEREFDIYWEYE			
LAGITKTVIECKDYNSKISIEKIDALSERKTRYSRFKTHICNQSRLSKWR"			
ORIGIN			
Query Match 3.1%; Score 44.6; DB 1; Length 5039;			
Best Local Similarity 46.1%; Pred. No. 6.2;			
Matches 149; Conservative 0; Mismatches 174; Indels 0; Gaps 0;			

QY 387 GTTCGATCTTTGTAATGCTATCTCCCTACGGAAGCTTCTTCAGATTGCTGGCGG 446  
Db 1501 GTACACACATTTGTAATTAATCCCTTACAAAAATAACAATCTGCAAGTATTACCG 1560  
QY 447 GCTTGCACACAGCGGTCGTCGATGTCGACGAACTTTACGTGGCTTCTGGTGGC 506  
Db 1561 TAAATATTAAGTCTATCGGTATGAGCTGTTAATCTATATCTGCTGTTTGTGCTCT 1620  
QY 507 AGCAGTCAATTTCCGTCACAAAGACGCGCGGGGTTTTCATGTTCTCTGATCTGGC 566  
Db 1621 TGCTATCATGCAATTAATAAACHAGGTGAATGCTAGCAATAATCCAGATCTTTTG 1680  
QY 567 GAACGGCCCTTACTATCTCAATTCGCCATTTGCTGATGACCGCGGTGAAGTCTCGATAT 626  
Db 1681 TAATGGCCCATATTAATCTACCTTCAGGAATTTTATATTTTTCAGCACTGTGCCATCAACA 1740  
QY 627 ACTTCATGTTGTCGAAAGTAGAACCAAGATTTTGGCGACGACGAAGTAAGCAAGAA 686  
Db 1741 TGTTCATATATTCGATAGTGAAGTCAATGCTCTGGAAGATGATGCTACTCCAAGAAA 1800  
QY 687 TGTTCATGTTGCTTTCAGTGTGA 709  
Db 1801 TATCATTAATTCATTAGTTAAGA 1823

## RESULT 9

AKI08161/c

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) cDNA clone:002-139-G12, full

insert sequence.

AKI08161

AKI08161.1 GI:32993370

FULL DNA; oligo capping.

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

AUTHORS

The Rice Full-length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-length cDNA Project Team,

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H.,

Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

Ohtsuki, K., Shishiki, T., Foundation of Advancement of International

Science Genome Sequencing &amp; Analysis Group: Otono, Y., Murakami, K.,

Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,

Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,

Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J.,

Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,

Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,

Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,

Hara, A., Hashidume, W., Hayashida, H., Imotani, K., Ishii, Y., Itoh, M.,

Kagawa, T., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,

Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Yoshino, M. and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from

japonica rice

Science 301 (5631), 376-379 (2003)

22752273

12869764

2 (bases 1 to 1164)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,

Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,

Hori, F., Hotta, J., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,

Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,

Kagawana, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,

Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,

Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Kusumegi, T., Li, C., Lu, M.,

Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,

Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,

Namiki, T., Narikawa, R., Niikura, J., Nishi, K., Nomura, K.,  
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,  
Osato, N., Ota, Y., Otono, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,  
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,  
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,  
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,  
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,  
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,  
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and  
Yoshimura, A.

## Direct Submission

## TITLE

## JOURNAL

Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of  
Agrobiological Sciences, Department of Molecular Genetics, Head of  
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki  
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,  
tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica  
rice.

URL : <http://cdna01.dna.affrc.go.jp/cDNA/>

NTAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,  
Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,  
Shikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,  
Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and  
Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otono, Y., Iida, Y.,  
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,  
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,  
Mizuno, K., Narikawa, R., Niikura, J., Oka, M., Ryu, R., Sugano, S.,  
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,  
Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center  
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,  
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,  
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,  
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,  
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,  
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,

Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,  
Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,  
Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,  
Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,  
Yasunishi, A. and Hayashizaki, Y.

## FEATURES

## source

1..1164

/organism="Oryza sativa (japonica cultivar-group)"

/mol type="mRNA"

/cultivar="Nipponbare"

/db xref="taxon:39947"

/clone="002-139-G12"

## ORIGIN

Query Match 3.0%; Score 44.2; DB 8; Length 1164;

Best Local Similarity 54.7%; Pred. No. 6.9;

Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 136 GATGATCTCCGGGAACGGTGGGTGCTCGACCCAGGAGCTGGTGGGTCTTGACC 195

Db 830 GAGGAGCGCGGCGTTTGAGGAGGGTGTGAAGACCCAGGCGCGATCGCGTGCACCGCC 771

QY 196 GCTGCGCTCGTCATCGGCTGCATCTGAACGCCCGAGCTGGTCCGCTCAGTAGTGCC 255

Db 770 GCGGCGAGCGCGCCCGAAGTGCATCCACGACGCGCGAGCGGAGGAGGTGACCC 711

QY 256 GTGGAACACCGACCCCTTTTCTGCTGCTTACCTTCGCGGCCAC 296

Db 710 GAGGAGCGCGCATCTGCACGCGGAGCGCGCGGCGGCCAC 670

## RESULT 10

CNS08CAM

LOCUS

CNS08CAM 146585 bp DNA linear PLN 21-NOV-2003

```
DEFINITION      Oryza sativa chromosome 12, . BAC OJ1268 D02 of library Monsanto
                  from chromosome 12 of cultivar Nipponbare of spp. japonica of Oryza
                  sativa (rice), complete sequence.
ACCESSION        AL831809
VERSION           AL831809.3  GI:25900557
SOURCE            HTG.
ORGANISM          Oryza sativa (japonica cultivar-group)
                  Oryza sativa (japonica cultivar-group)
                  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                  Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE         1 (bases 1 to 146585)
                  Choinsne,N., Orjeda,G., Cattolico,L., Demange,N., Wincker,P.,
                  Seguren,B., Pelletier,E., Scarpelli,C., Salanoubat,M.,
                  Weissenbach,J. and Quetier,F.
                  Oryza sativa chromosome 12 sequencing
                  Unpublished
                  2 (bases 1 to 146585)
                  Genoscope.
                  Direct Submission
                  Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
                  BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
                  - Web : www.genoscope.cns.fr)
                  On Nov 28, 2002 this sequence version replaced gi:24817645.
                  Center: Genoscope / Centre National de Sequencage
                  Web site: http://www.genoscope.cns.fr/
                  Contact: Segref@genoscope.cns.fr
                  -----
                  The following sequence is oriented from the T7 to the SP6 end. The
                  nucleotide sequence of this BAC clone was generated by combining
                  Monsanto and Genoscope sequencing data.
                  Upstream BAC (overlapping the T7 end) : OJ1559_C07 (AC=AL731887)
                  Downstream BAC (overlapping the SP6 end) : OJ1119_E02 (AC=AL731762)
                  ----- Finishing boundaries
                  FINISHED SEGMENT STARTS AT BASE 1
                  FINISHED SEGMENT ENDS AT BASE 146585
                  -----
FEATURES          Location/Qualifiers
                  1..146585
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /sub_species="japonica"
                     /db_xref="taxon:39947"
                     /chromosome="12"
                     /clone="OJ1268_D02"
                     /clone_lib="Monsanto"
ORIGIN
Query Match      3.0%; Score 44.2; DB 8; Length 146585;
Best Local Similarity 54.7%; Pred No.10;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
QY 136 GATGATCTCCGCGGAACGGTGCGGGTGCTGACCCAGGAGCTGGTGTGCGGTCTCTGACC 195
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99220 GAGGAGCGCGGCTTGGAGGAGGTGTGAAGACCCAGGCGCGATGCCGTGACCCGCC 99279
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 196 GCTGCGCTCTGATCGGCTGCATCTGACCGCCGACGCTTCCGCTCCACGTAGTTGCC 255
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99280 GCGGCGAGCGCGCGCGAAGTGCATCCACAGCAGCGCGAGTGGCAGAGGTTGCACCC 99339
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 256 GTGGAAACCGACCCCTTTTCGTGCTTACCTGCGGCCAC 296
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99340 GAGGAGCGCGCACTGCGAGCGCGAGCGCGCGGGGGCGCCAC 99380
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 11
AP005939/c
LOCUS             300900 bp DNA linear BCT 12-SEP-2003
DEFINITION        Bradyrhizobium japonicum USDA 110 DNA, complete genome, section
                    5/31.
ACCESSION         AP005939 BA000040
```

```
VERSION          AP005939.1  GI:27349336
KEYWORDS          Bradyrhizobium japonicum USDA 110
SOURCE            Bradyrhizobium japonicum USDA 110
ORGANISM          Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
                  Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE         1
                  Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiimi,T.,
                  Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,
                  Kohara,M., Matsumoto,M., Shimpō,S., Tsuruoka,H., Wada,T., Yamada,M.
                  and Tabata,S.
                  Complete genomic sequence of nitrogen-fixing symbiotic bacterium
                  Bradyrhizobium japonicum USDA110
                  DNA Res. 9 (6), 189-197 (2002)
                  22484998
                  12597275
                  2
                  Kaneko,T., Nakamura,Y., Sato,S., Minamisawa,K., Uchiimi,T.,
                  Sasamoto,S., Watanabe,A., Idesawa,K., Iriguchi,M., Kawashima,K.,
                  Kohara,M., Matsumoto,M., Shimpō,S., Tsuruoka,H., Wada,T., Yamada,M.
                  and Tabata,S.
                  Complete genomic sequence of nitrogen-fixing symbiotic bacterium
                  Bradyrhizobium japonicum USDA110 (supplement)
                  DNA Res. 9 (6), 225-256 (2002)
                  22485002
                  12597279
                  3 (bases 1 to 300900)
                  Kaneko,T.
                  Direct Submission
                  Submitted (20-NOV-2002) Takakazu Kaneko, Kazusa DNA Research
                  Institute, The First Laboratory for Plant Gene Research; 2-6-7
                  Kazusa-kanatari, Kisarazu, Chiba 292-0812, Japan
                  (E-mail:kaneko@kazusa.or.jp,
                  URL:http://www.kazusa.or.jp/rhizobase/,
                  Tel:81-438-52-3935 (ex.2338), Fax:81-438-52-3934)
FEATURES          Location/Qualifiers
                  1..300900
                     /organism="Bradyrhizobium japonicum USDA 110"
                     /mol_type="genomic DNA"
                     /strains="USDA110"
                     /db_xref="taxon:224911"
                     103..756
                     /gene="blr1086"
                     103..756
                     /gene="blr1086"
                     /note="ORF ID:blr1086
                     TetR family"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulatory protein"
                     /protein_id="BAC46351.1"
                     /db_xref="GI:27349337"
                     /translation="MKSPDETTLDTNTVYRIILMPKRSKLDAILDAGLKVMFRITGY
                     HGTSVRDYTAAGAPQGSFTNHFPSKEAFSEVLDYFDVTFGLVAEALDITLTPRA
                     RLKRLDITIGRLDAGTGRCGLIGLSLEATGSGSEMRLTRLTETFAEWRVPPACIA
                     EAQKGEITASPEPGLADFLASWQGAILLRMKVDNRNPKALERFKTIAPQTVFKELT"
                     753..1685
                     /gene="blr1087"
                     753..1685
                     /gene="blr1087"
                     /note="ORF ID:blr1087"
                     /codon_start=1
                     /transl_table=11
                     /product="haloalkane dehalogenase"
                     /protein_id="BAC46352.1"
                     /db_xref="GI:27349338"
                     /translation="MSKRIEIRRAPVLGSSMAYRETCAODAPVVLFLHGNPTSSHI
                     WRNLLPLSPVAHCIAFDLIGFGQSGKFDIATFFDHYRLDAFIEQRGVTSAYLVAQ
                     DMGTALAFHLAARFPDFVRGLAFMEFIRPMTWQDFHTEVAEBQDHAAARAFKRF
                     RTPGGEAMILEANAFVERVLPGGIVRKIGDEMAPYRTPPTPSRRVLPAPFRLP
                     IAGEPADVVEALQSAHALAASSYPKLLFTGEPGALVSPFEAFERFAASITRCALIRLG
                     ACLHYLDQHDHDAIGRSVAGTWAGIEAVRPQLAA"
                     1941..2372
```

```
/gene="blrl1088"
1941..2372
/gene="blrl1088"
/notes="ORF ID:blrl1088
putative GNAAT family acetyltransferase"
/codon_start=1
/transl_table=11
/protein_id="BAC46353.1"
/db_xref="GI:27349339"
/translation="MPVWNEDPAAIDWNLSALYRTAPLGNKTPADLALVFGNSMFC
AFAYDQRLVAGRALADGRDCAYLDCVAVRFDCCGGGLGREIIIRLLARCRGHRKII
LIYVPGKEGYERLGRFRMTTAMAFDDQASAHQRYLTAP"
2795..3850
/gene="blrl1089"
2795..3850
/gene="blrl1089"
/notes="ORF ID:blrl1089
hypothetical protein"
/codon_start=1
/transl_table=11
/protein_id="BAC46354.1"
/db_xref="GI:27349340"
/translation="MLBAIRRAMTFLRQKQVLHLGVVSVAVIGIACYVLYHMLRGI
DFNEVEIAKSTSPSQIAMAALFVAGYFTLFDYLFVRAIGHAHPYRINALAAFT
SYSIGHNVGASVFTGGAVRYRIYSAYGLNAIDVAKICFLAGLTFWLGNAAVLGLGISY
HPFAAASIDQLPWLNRLTALMIIVGLGVVWVWVQPRVVGPGPTVLPVGLPGPLTIL
QIAIGIIDLGFALAMYLVPDRPNLGFVVAVIVFSATLLGFASHRSGLGVFDDAM
LVGLWMDREELGMLLFRVLYLSPFVLSVILLIFREVIIGARSKRLQQAALKLDP
GPAREAAVYRERSDGTG"
3910..5208
/gene="phoR"
3910..5208
/gene="phoR"
/notes="ORF ID:blrl1090"
/codon_start=1
/transl_table=11
/product="phosphate regulon, two-component sensor
histidine kinase"
/protein_id="BAC46355.1"
/db_xref="GI:27349341"
/translation="MAIDAPSSPTAQPSWDRLRHSTIILIAAALSVVSLGELSVL
HAVAVFLCIAAALIPLRLHDTAASRDDVRRINPVESAAVAAGMPDPVAVLLDRAG
RVHLNAAQAALPALRNLAEQALRSPETIITALRESIATTEPRRATYLDHPVDRW
MELIITPVPTVFTGGADKMLTFHDQTPLRVVEBMRADPFVANASHELRTPLAALSG
FIDTLOQAKDDPKARERFLIGHNOATRMARLIDLLSLRVELSAHVLRPDTLLD
PIIROVADGLEPLAREQVEVEHLPEPPVMIAGDBELLRLFENLLENALKYGSAGG
RVIVTSQAATGTQEURVWADFPGIAPHELPRLTERFYKRVVDVGDRLQGGTGIG
LSLVHLNRHRRGRLLIESVPKQGAFTTACFPQIRPVTV"
5341..6366
/gene="ptsS"
5341..6366
/gene="ptsS"
/notes="ORF ID:blrl1091"
/codon_start=1
/transl_table=11
/product="ABC transporter phosphate-binding protein"
/protein_id="BAC46356.1"
/db_xref="GI:27349342"
/translation="METSNNFLKTIIVAAGLVAASTTAFAADITGAGATFPFPFVSKW
ADAYKKTGNGLNYQSIGSGGKIQIQAKTVTTFGASDAPLKAOLEKDLGVQPMVMG
ALVPVNIIEGVKPEGMVFDGETIASIYLGKTIKWDDAAIKLNLNKLPSDAITVVR
SDSGTTFNFTNVLKASADWKSIVGEGTAVEMVPGVAKNGEVSGNIGOTKNISYI
VEYAVAKNKLTVTLGVNKAQKVPQPTVEAFQAAASNADWAKARGVYVILTPQPEKS
WPIATAATEFILMHKDATDKAASQEAIKFFFWAFNKGSKAAELDIYIPMEGVVTLIEKT
WAAEIKS"
6487..7485
/gene="patC"
6487..7485
/gene="patC"
/notes="ORF ID:blrl1092"
/codon_start=1
/product="phosphate ABC transporter permease protein"
/protein_id="BAC46357.1"
/db_xref="GI:27349343"
/translation="MAEMAVQSDVIDDAGPYDRAKALSAFKLGDVTFYWTIRLSAIV
LLILGGIIVSLIIVGAFPAIKEYGFSFLMTWAPSDAPPVLGALGPMYGTIVTFSFIAM
LIAIPVGLGIAIFLTELCPQMLRRPILGAIELLAIGPSIIYGMGFGVILGFLANTQ
PFMIKIPDGVPLGAIAPAGPSYLSFNAALILAIIMLPFTITSIVSVKTPVPVLKE
AAYGVGCTTWEVSRSVVIPTRVGIVGVMLALGRALGETMAVTFIIGNSFRISSSF
APGTTISAAIASFEASDGLHQSLGILLGLLLVLTFVFLVLAARMLMLRLEKAGN"
7490..8335
/gene="pstA"
7490..8335
/gene="pstA"
/notes="ORF ID:blrl1093"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter permease protein"
/protein_id="BAC46358.1"
/db_xref="GI:27349344"
/translation="MNPVYSRRRKDIIVVGLCIAAAGVTWMLALITLLVNGIAG
LNLEIFVAUTPPPGSTEGGLRNAIVGSIMTVLGVGIGAGLGLPAGTYLAEGNRDKL
TSVIRINDILLSPSIIIGLFIVGAVVPMRGFSATAGSLALAVIIVPVLRTDM
LLVLPNALREASAGLGLPSLVIKRIAYRAARSLGITVLLATARVAGETAPLLFTAL
SNOFFSLGLNKTWANLPVTINNFPVQSPYAYWKQLAWSGALLITITVLALNIGARILGA
ERTAK"
8332..9153
/gene="pstB"
8332..9153
/gene="pstB"
/notes="ORF ID:blrl1094"
/codon_start=1
/transl_table=11
/product="phosphate ABC transporter ATP-binding protein"
/protein_id="BAC46359.1"
/db_xref="GI:27349345"
/translation="MSELSVSMSAAGLPQAPLLPEAPAKVTVRNLFNFYGEHALKN
INLTGNRVTAFTFGSGGCKTLRIIFNRMYDLYPGQRTGOLMDOTNILDGKLDL
NLLRARVGVFOKPTFPFPIYENIAFGIRLYEKISKSEMDRVEKALRGALWNEVK
DKLNASGLSLSGQQORLCIARTVAVRPEVILDEPCSLDPISTAKVELIQLSEN
YTIIVHNHQARVSDKTAFFMYLGELIEFDFTSKIFTSPSRRTQDYITGRFG"
9165..9881
/gene="phoU"
9165..9881
/gene="phoU"
/notes="ORF ID:blrl1095"
/codon_start=1
/transl_table=11
/product="phosphate transport system regulatory protein"
/protein_id="BAC46360.1"
/db_xref="GI:27349346"
/translation="MGSEHTAKAFDLDLQELTRLVAEMGGIAERMIVDSVALIRRDV
```

Query Match 2.9%; Score 43; DB 1; Length 300900;  
Best Local Similarity 49.3%; Pred. No. 22;  
Matches 112; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY	125	TCGTTCCGGTCCGATGATCTCCCGGNAACGGTGGCGGTGCTGACCCAGGAGCTGGTGTGCG	184
Db	32318	TGATTGAAGCCGAAGATCGCTCGGCATGCTCGCGCAGCCCTCGAGAGCCGCGACGATC	32259

QY	185	GGTCTCTGACCGGTCCGCTCGATCGGCTGCATCTCACTCAACGCCGCCAGCTTGGCGTCC	244
Db	32258	GGGACCTTCCCGGAGCGGTGACCGACCCGCCGCCACCGCGCGCGCGGTGGCGCCC	32199

QY	245	ACGTAGTTCGCGTGAACCCGACCCCTTTGTCGTGCTTTTACCTGCGCGCCACCTCGAGG	304
Db	32198	GCCGCGATGTTGTCGATGCCACATAATGTTGTCGCGCGGACGACGCGACGTC	32139

QY	305	AATGTGCGNAACCTTACGGCATCTCTTACGACCTTAGTCGAGGGCGAC	351
Db	32138	AGCGTCACCAACCTTGGTGGCGGCATCGACGAGATCGTTGATCGCGGC	32092

RESULT 12



RESULT 13	AF246703/c	21376 bp	DNA	linear	STN 06-NOV-2000
LOCUS	AF246703				
DEFINITION	Synthetic construct				
ACCESSION	AF246703				
VERSION	AF246703.2	GI:11079645			
KEYWORDS					
SOURCE	synthetic construct				
ORGANISM	synthetic construct				
REFERENCE	1 (bases 1 to 21376)				
AUTHORS	Bascones, E., Imperial, J., Ruiz-Argueso, T. and Palacios, J.M.				
TITLE	Generation of new hydrogen-recycling Rhizobiaceae strains by introduction of a novel hupD minitransposon				
JOURNAL	Appl. Environ. Microbiol. 66 (10), 4292-4299 (2000)				
MEDLINE	20466836				
PUBMED	11010872				
REFERENCE	2 (bases 1 to 21376)				
AUTHORS	Bascones, E., Imperial, J., Ruiz-Argueso, T. and Palacios, J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (17-MAR-2000) Biotechnology, Univ. Politecnica de Madrid, Ciudad Universitaria s/n, Madrid, Madrid 28040, Spain				
COMMENT	On Nov 3, 2000 this sequence version replaced gi:11034796.				
FEATURES	Location/Qualifiers				
source	1..21376				
	/organism="synthetic construct"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:32630"				
	/transposon="TnH100"				
	/note="miniTn5-type transposon"				
repeat_region	1..19				
	/note="Tns left repeat"				
	/rpt_type=inverted				
misc_feature	complement(28..2063)				
	/note="Omega interposon"				
misc_feature	2224..3257				
	/note="corresponds to the Rhizobium leguminosarum bv. viciae putative 2-hydroxychromene-2-carboxylate isomerase and putative alcohol dehydrogenase genes in GenBank Accession Number U23040"				
misc_feature	3258..3383				
	/note="corresponds to the Rhizobium leguminosarum bv. viciae hupSL gene promoter in GenBank Accession Number U36924"				
misc_feature	3384..19767				
	/note="corresponds to Rhizobium leguminosarum sequence in GenBank Accession Number XS2974"				
repeat_region	21358..21376				
	/note="Tns right repeat"				
	/rpt_type=inverted				
ORIGIN					
Query Match	2.9%; Score 42.8; DB 12; Length 21376;				
Best Local Similarity	58.7%; Pred. No. 20;				
Matches	74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;				
QY	178	GGTGTGGGCTCTGACCGTGCCTCGTCGATCGGCTGCATCTAGAACGCCGACGTT	237		
Db	14029	GGCAATGGCTTCTGTGTCGCGGCTTGGCAGCGCCGACAGGCGGATCGCCATCGTT	13970		
QY	238	CGCGTCCACGTAGTTGCGGTGGAACCGACCCCTTTTGTGTCGTCTTACCTGCGCGCCACC	297		
Db	13969	CGCGGCGTGCAGGAAGTGAACCGCGACGCCCTTGAGCGCAACGATCGCGCGCC	13910		
QY	298	CTCGAG 303			
Db	13909	TTGAG 13904			
RESULT 14	AC108223/c	175376 bp	DNA	linear	HTG 28-FEB-2003
LOCUS	AC108223				
DEFINITION	Oryza sativa (japonica cultivar-group chromosome 11 clone				

mat_peptide	1262..3004				
	/gene="hupL"				
	/product="hydrogenase large subunit"				
	/function="hydrogenase large subunit"				
	/citation={6}				
evidence-experimental	1263..1274				
RBS	/gene="hupL"				
RBS	3053..3058				
gene	3066..3785				
	/gene="hupC"				
CDS	3066..3785				
	/gene="hupC"				
	/function="hydrogenase accessory protein"				
	/note="Integral membrane protein"				
	/codon start=1				
	/transl_table=11				
	/protein_id="CAA37150.1"				
	/db_xref="GI:48723"				
	/db_xref="GOA:P27648"				
	/db_xref="SWISS-PROT:P27648"				
	/translation="MTIHTLAADAHGEKAIERQSVYVYAPVRIWHWLNFSILTLA				
	LTCGFTGSLPSVPGRAANFLMGVIRFIHFAGQLLAVFLIRVYVAEYVGNVHAROI				
	FVYVFGSGRPEKWEHVEGMYTFLVQPKKYVGHNFELPQFTFLPFLTLLFWAITGF				
	ALYSEAGRDSWYSYSLFGWVFSIWPNSQDIIHTHLLGMWVILFVNVVHIYVAVREDDIM				
	SQSIISMSIGSERLFPKRED"				
RBS	3076..3081				
	/gene="hupC"				
gene	3786..4394				
	/gene="hupD"				
CDS	3786..4394				
	/gene="hupD"				
	/function="putative specific protease involved in HupL				
	maturaton"				
	/codon start=1				
	/transl_table=11				
	/protein_id="CAA37151.1"				
Query Match	2.9%; Score 42.8; DB 1; Length 17786;				
Best Local Similarity	58.7%; Pred. No. 19;				
Matches	74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;				





**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: October 1, 2004, 19:21:23 ; Search time 625.714 Seconds  
(without alignments)  
9919.257 Million cell updates/sec

Title: US-10-668-047-1  
Perfect score: 1461  
Sequence: 1 gtgcattccgacgcagcac.....ttccggaatcgtcgtga 1461

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	49.6	3.4	2000	7	ADA71938	Ada71938 Rice gene
2	45	3.1	110000	2	AAV21209_04	Continuation (5 of
3	44.8	3.1	536	9	ADB68842	Adb68842 Minority
4	42.2	2.9	1947	6	AAL40127	Aal40127 Isoprenoi
5	41.2	2.8	1425	7	ACA26157	Aca26157 Prokaryot
6	41.2	2.8	113193	7	RAA54645	Raa54645 Streptomy
7	41	2.8	1311	7	ACA26275	Aca26275 Prokaryot
8	39.6	2.7	1503	7	ACA25789	Aca25789 Prokaryot
9	39.6	2.7	2000	7	ADA71938	Ada71938 Rice gene
10	39.4	2.7	987	7	ACA43826	Aca43826 Prokaryot
11	39.4	2.7	1119	7	ACA23836	Aca23836 Prokaryot
12	39	2.7	975	7	ADA69358	Ada69358 Rice gene
13	38.8	2.7	551	7	ABX74255	Abx74255 Rice gene
14	38.8	2.7	9810	2	AAZ32025	Aaz32025 Human MET
15	38.8	2.7	9810	5	AAC90082	Aac90082 AF018073
16	38	2.6	2000	9	ADC08454	Adc08454 Rice DNA
17	37.6	2.6	633	8	ADA49246	Ada49246 Maize gen
18	37.6	2.6	110000	2	AZ01425_01	Continuation (2 of
19	37.6	2.6	110000	2	AZ01425_02	Continuation (3 of
20	37.4	2.6	789	6	ABN87153	Abn87153 Lolium pe
21	37.2	2.5	687	7	ACA38429	Aca38429 Prokaryot
22	37.2	2.5	690	7	ACA40639	Aca40639 Prokaryot
23	37.2	2.5	110000	4	AAI99682_26	Continuation (27 o

24	37.2	2.5	110000	4	AAI99683_25	Continuation (26 o
25	37.2	2.5	110000	4	AAI99683_26	Continuation (27 o
26	36.8	2.5	925	2	AAV15076	Aav15076 Hybrid DN
27	36.8	2.5	38726	4	AAV59513	Aav59513 Propionib
28	36.8	2.5	38726	7	ACF64442	Acf64442 Propionib
29	36.6	2.5	367	3	AAC56312	Aac56312 Pinus rad
30	36.6	2.5	3453	7	ACA26750	Aca26750 Prokaryot
31	36.4	2.5	1017	9	ADC23900	Adc23900 DNA seque
32	36.4	2.5	2522	6	ABK98456	Abk98456 Consensus
33	36.4	2.5	6291	4	ABL15665	Ab115665 Drosophil
34	36.4	2.5	14260	4	ABL15664	Ab115664 Drosophil
35	36.4	2.5	14615	4	AAS59577	Aas59577 Propionib
36	36.4	2.5	14615	7	ACF64506	Acf64506 Propionib
37	36.4	2.5	30001	2	AA761016	Aat61016 Total DNA
38	36.4	2.5	30001	2	RAA05110	Raa05110 S. aureof
39	36.2	2.5	788	6	ABN87154	Abn87154 Lolium pe
40	36.2	2.5	1233	7	ACA42148	Aca42148 Prokaryot
41	36.2	2.5	1395	4	AAS51511	Aas51511 Pseudomon
42	36.2	2.5	1395	7	ACA19477	Aca19477 Prokaryot
43	36.2	2.5	1486	6	ABN87145	Abn87145 Lolium pe
44	36.2	2.5	2277	3	AAZ53328	Aaz53328 Neisseria
45	36.2	2.5	2295	3	AAZ53319	Aaz53319 Neisseria

ALIGNMENTS

RESULT 1  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
XX  
AC ADA71938;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 5263.  
XX  
KW Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX  
OS Oryza sativa.  
XX  
PN WO2003000898-A1.  
XX  
PD 03-JAN-2003.  
XX  
PF 22-JUN-2001; 2001WO-IB001105.  
XX  
PR 22-JUN-2001; 2001WO-IB001105.  
XX  
(SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX  
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX  
DR WPI; 2003-175290/17.  
XX  
PT Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX  
PS Claim 27; SEQ ID NO 5263; 899pp; English.  
XX

The present invention relates to a method (M1) for identifying genes involved in plant resistance or response to pathogenic infection. M1 comprises identifying a gene whose expression is significantly altered in the incompatible interaction of plant gene expression relative to the expression of the gene in an uninfected plant, in a mutant plant that does not express a gene associated with response to pathogenic infection, or in a corresponding incompatible or compatible interaction. (M1) is useful for conferring resistance to resistance or tolerance to a plant to bacterial, fungal or viral infection. The present sequence was used to









Qy	168	CCCAGGAGCTGGTGTGCGGCTCTGTACCGCTGCCCTCGTCGATCGGCTGCATCACTGAACG	227
Db	756	CGGCGTCTGTGGCGCAGCTCGATTCGATACGCACATCTGTGCACCACTTCTCTCGTTTCGC	815
Qy	228	CCCCGAGCTTCGGGTCCACGTAGTTGCCGTGAAACCGACCCCTTTTGTTCGTGCCCTTACCT	287
Db	816	GCA CGGCGACCGCAGCCGACGGAGCGCTGCCGTCGACCAAGGCTGGCGACGGGATCGC	875
Qy	288	CGCGGCCACCCCTGGAGGAATGTTCGAAAGCGCTTACGGGCATCTCCTACGACTAGTCGAGG	347
Db	876	GCGCACTTACCGCGGGTGTCCCCGAA CGCGCCACCGTGGCGACCGAGCTCGCGGCCG	935
Qy	348	CGACTATTTCCTTACCAAGGGGCCAAGCTGA	380
Db	936	GCCGGGCTTTTCGCTGCGCGGCGGCGACGCTCGA	968

## RESULT 8

ACA25789  
ID ACA25789 standard: DNA: 1503 BP.

XX ACA25789:

AA	19-JUN-2003	(first entry)
DT		

XX DE Prokaryotic essential gene #7446.

AA Antisense; ds, prokaryotic essential gene; cell proliferation;  
KW KW drug design; gene.

OS Burkholderia fungorum.

AA  
PN  
WO200277183-A2.

03-OCT-2002.

21-MAR-2002: 2002WO-US009107.

21-MAR-2001: 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX  
PA (E.I.TT-) E.I.TTRA PHARM INC.

XX	Wang L,	Zamudio C,	Malone C,	Haselbeck R,	Ohlsen KL,	Zyskind JW;
PI	wall D,	Trawick JD,	Carr GJ,	Yamamoto R,	Forsyth RA,	Xu HH;

XX  
DP  
WPT: 2003-029926/02.

DR WFI; Z003=029928/  
DR P-PSDB: ABU21919.

New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX  
PS Claim 14: SEO ID NO 13659; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological



XX AC A23836;  
 XX 19-JUN-2003 (first entry)  
 XX DE Prokaryotic essential gene #5493.  
 XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 XX KW drug design; gene.  
 XX OS Borrelia cepacia.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELITRA PHARM INC.  
 XX PW Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;  
 XX P-PSDB; ABU19966.  
 XX WPI; 2003-029926/02.  
 XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 XX PT for homologous nucleic acids required for cellular proliferation to  
 XX PT isolate candidate molecules for rational drug discovery programs.  
 XX PS Claim 14; SEQ ID NO 11706; 1766pp; English.  
 XX CC The invention relates to an isolated nucleic acid comprising any one of  
 XX CC the 6213 antisense sequences given in the specification where expression  
 XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 XX CC encoding a polypeptide whose expression is inhibited by the antisense  
 XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 XX CC polypeptide or its fragment whose expression is inhibited by the  
 XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 XX CC proliferation or the activity of a gene in an operon required for  
 XX CC proliferation; (7) identifying a compound that influences the activity of  
 XX CC the gene product or that has an activity against a biological pathway  
 XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
 XX CC identifying a gene required for cellular proliferation or the biological  
 XX CC pathway in which a proliferation-required gene or its gene product lies  
 XX CC or a gene on which the test compound that inhibits proliferation of an  
 XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 XX CC compound's activity; (11) a culture comprising strains in which the gene  
 XX CC product is overexpressed or underexpressed; (12) determining the extent  
 XX CC to which each of the strains is present in a culture or collection of  
 XX CC strains; or (13) identifying the target of a compound that inhibits the  
 XX CC proliferation of an organism. The antisense nucleic acids are useful for  
 XX CC identifying proteins or screening for homologous nucleic acids required  
 XX CC for cellular proliferation to isolate candidate molecules for rational  
 XX CC drug discovery programs, or for screening homologous nucleic acids  
 XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 XX CC prokaryotic essential genes. Note: The sequence data for this patent did  
 XX CC not form part of the printed specification, but was obtained in  
 XX CC electronic format directly from WIPO at  
 XX CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 1119 BP; 221 A; 382 C; 360 G; 156 T; 0 U; 0 Other;  
 Query Match 2.7%; Score 39.4; DB 7; Length 1119;

Best Local Similarity 48.1%; Pred. No. 0.38;  
 Matches 112; Conservative 0; Mismatches 121; Indels 0; Gaps 0;  
 QY 90 CACTCCCATGAAGCGCGCCACGCTGATGGCTTCATGTTGGGTGATGATCTCGCGG 149  
 DB 862 CACCCGCGAAGCGACCGCTGTCGATACGCGCCCGTCCCTTCGCGACCGGAATCGTGT 803  
 QY 150 AACGGTGGGGTCTCGACCCAGGAGCTGGTGTCCGGGTCTCTGACCGCTGCCCTCGTCA 209  
 DB 802 GCGCAGCAGGCTGTCGACGAGCCGCTGTCGTCCTTGAACGCGAGTAGCCCGGTAGA 743  
 QY 210 TCGGCTGCATGTAACGCGCCGACGTTGCGGTCCACGTAGTTCGCGTGGAAACCGACCC 269  
 DB 742 TCTTCGCTTGTCTGAAACGCTACGATGTTCCGCTGGTAGATCGTCTCTTCGCGCGCG 683  
 QY 270 CTTTGTGTCCTTACCTGCGCGCCACCTCGAGGAATGTTCGGAACGCTTACG 322  
 DB 682 AGTTGTGCTGTCTGCTGCGCGCGCGCGATGCTGAGCGGACCCATCAG 630

RESULT 12  
 ADA69358/c  
 ID ADA69358 standard; DNA; 975 BP.

XX AC ADA69358;  
 XX 20-NOV-2003 (first entry)  
 XX DE Rice gene, SEQ ID 2681.  
 XX KW Plant; bacterial infection; fungal infection; viral infection; rice;  
 XX KW gene; ds.  
 XX OS Oryza sativa.  
 XX PN WO2003000898-A1.  
 XX PD 03-JAN-2003.  
 XX PF 22-JUN-2001; 2001WO-IB001105.  
 XX PR 22-JUN-2001; 2001WO-IB001105.  
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
 XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
 XX WPI; 2003-175290/17.

XX PT Identifying at least one gene involved in plant resistance or response to  
 XX PT pathogenic infection for conferring resistance or tolerance to a plant to  
 XX PT bacterial, fungal or viral infection by determining or detecting plant  
 XX PT gene expression.  
 XX PS Claim 6; SEQ ID NO 2681; 899pp; English.  
 XX CC The present invention relates to a method (M1) for identifying genes  
 XX CC involved in plant resistance or response to pathogenic infection. M1  
 XX CC comprises identifying a gene whose expression is significantly altered in  
 XX CC the incompatible interaction of plant gene expression relative to  
 XX CC expression of the gene in an uninfected plant, in a mutant plant that  
 XX CC does not express a gene associated with response to pathogenic infection,  
 XX CC or in a corresponding incompatible or compatible interaction. (M1) is  
 XX CC useful for conferring resistance to resistance or tolerance to a plant to  
 XX CC bacterial, fungal or viral infection. The present sequence was used to  
 XX CC illustrate the invention.

XX SQ Sequence 975 BP; 182 A; 364 C; 284 G; 145 T; 0 U; 0 Other;  
 Query Match 2.7%; Score 39; DB 7; Length 975;  
 Best Local Similarity 53.6%; Pred. No. 0.47;  
 Matches 81; Conservative 0; Mismatches 70; Indels 0; Gaps 0;





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 23:40:18 ; Search time 116.496 Seconds  
(without alignments)  
6959.771 Million cell updates/sec

Title: US-10-668-047-1

Perfect score: 1461

Sequence: 1 gtgcattccgacgcagcac.....ttccggaatcgtcgtga 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*

2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*

3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*

4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*

5: /cgn2\_6/ptodata/2/ina/PTUS\_COMB.seq.\*

6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	3.1	1664976	4	US-08-916-421B-1
2	41.2	2.8	840	4	US-09-252-991A-10289
3	41.2	2.8	1260	4	US-09-252-991A-10493
C 4	41.2	2.8	1635	4	US-09-252-991A-10567
C 5	41.2	2.8	1743	4	US-09-252-991A-10674
6	38.8	2.7	551	4	US-09-470-191-27
7	38	2.6	507	4	US-09-252-991A-15931
8	38	2.6	1056	4	US-09-252-991A-15895
9	38	2.6	3357	4	US-09-252-991A-15868
C 10	37.8	2.6	1257	4	US-09-252-991A-4342
C 11	37.8	2.6	1353	4	US-09-252-991A-4513
C 12	37.8	2.6	1419	4	US-09-252-991A-15497
C 13	37.8	2.6	2016	4	US-09-252-991A-15291
C 14	37.8	2.6	2508	4	US-09-252-991A-15550
C 15	37.4	2.6	930	4	US-09-252-991A-4633
16	37.4	2.6	1179	4	US-09-252-991A-4241
C 17	37.4	2.6	1719	4	US-09-252-991A-4534
18	37.2	2.5	4403765	3	US-09-103-840A-2
19	37.2	2.5	4411529	3	US-09-103-840A-1
20	36.8	2.5	525	4	US-09-252-991A-14501
C 21	36.8	2.5	922	3	US-09-189-060B-73
C 22	36.8	2.5	1083	4	US-09-252-991A-3016
C 23	36.8	2.5	1092	4	US-09-252-991A-14759
24	36.8	2.5	1452	4	US-09-252-991A-2653
25	36.8	2.5	2145	4	US-09-252-991A-14638
C 26	36.8	2.5	2487	4	US-09-252-991A-14985
C 27	36.4	2.5	30001	1	US-08-125-468-1

C 28	36.4	2.5	30001	2	US-08-474-933-1	Sequence 1, Appli
C 29	36.2	2.5	1329	4	US-09-252-991A-15277	Sequence 15277, A
C 30	36.2	2.5	1551	4	US-09-252-991A-15327	Sequence 15327, A
31	36	2.5	1098	4	US-03-252-991A-4322	Sequence 4322, Ap
C 32	36	2.5	1137	4	US-09-252-991A-4155	Sequence 4155, Ap
C 33	36	2.5	3924	4	US-09-252-991A-4296	Sequence 4296, Ap
C 34	35.8	2.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 35	35.8	2.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 36	35.4	2.4	2634	3	US-08-941-936-1	Sequence 1, Appli
C 37	35.4	2.4	3120	4	US-09-221-017B-54	Sequence 54, Appli
C 38	35	2.4	645	4	US-09-252-991A-2971	Sequence 2971, Ap
C 39	35	2.4	678	4	US-09-252-991A-893	Sequence 893, App
40	35	2.4	1272	4	US-09-252-991A-16045	Sequence 16045, A
C 41	35	2.4	1575	4	US-09-252-991A-1139	Sequence 1139, Ap
C 42	35	2.4	1587	4	US-09-023-655-1192	Sequence 1192, Ap
C 43	35	2.4	1830	4	US-09-252-991A-3164	Sequence 3164, Ap
C 44	35	2.4	1980	4	US-09-252-991A-16555	Sequence 16555, A
C 45	35	2.4	2859	4	US-09-252-991A-944	Sequence 944, App

#### ALIGNMENTS

#### RESULT 1

US-08-916-421B-1

; Sequence 1, Application US/08916421B

; Patent No. 6503729

; GENERAL INFORMATION:

; APPLICANT: Bult et al.

; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus

; Patent No. 6503729

; TITLE OF INVENTION: jannaschii

; FILE REFERENCE: PB275

; CURRENT APPLICATION NUMBER: US/08/916,421B

; PRIOR FILING DATE: 1997-08-22

; PRIOR APPLICATION NUMBER: US 60/024,428

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 1664976

; TYPE: DNA

; ORGANISM: Methanococcus jannaschii

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (28222)..(28222)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (28257)..(28258)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (84773)..(84773)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (84808)..(84808)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (84812)..(84812)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (98120)..(98120)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (98159)..(98159)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (98239)..(98239)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (98266)..(98266)

; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc feature

; LOCATION: (98343)..(98343)

```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191389)..(191389)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231380)..(231380)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

Query Match 3.1%; Score 45; DB 4; Length 1664976;

Best Local Similarity 47.4%; Pred. No. 0.058;

Matches 135; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```
QY 681 AGAGAAATGTCATCGTTGCTTTTCAGTGTGAGCGCGCAAGCTCTAGTGTGCTCTCTAG 740
Db 416035 AGTTAATGTTCTTATCTGGTGTGCATTTACAGACCAAGGATTTATCGAAGGCAAGCTTTA 416094
QY 741 GTCCGTCGCACATGGAGAGAGTCGATCGCAGTTCTGTGCGTTTCTGCGCTTCTTCA 800
Db 416095 TGAATTAGCCAAAGCTAAAAAGTTGCTTCCAAAGTAAGATAAAGTTTTTTCAGATTTCA 416154
QY 801 TGAATGAAGACGATGACAAAATCGTCACITTCGCGGAAAGCGCATCGGTGCGCGGCG 860
Db 416155 TGTAAAGCATGCATATCATTTTATAGACITTTGAAGCTCATTTGATACCGTTGAGAG 416214
QY 861 GAGGTTTACTCTCGCTGATCTCGGCATCGGTGTAAGTACGGAAAGTTGTTGATTTTCG 920
```

Db 416215 AGGTTAGCTGATGCTGTAATATATCAGCGGTAAAGAGAACGGGAAAGAGGTTGATATTGA 416274  
QY 921 CAATCGTCAGTATTGACCGGATAACCTGGATGCTTCACGGCTTGT 965  
Db 416275 AAAGCTAAATAGCTAAAGGAATTGGTTGATGTTCCAGTTATTGT 416319

## RESULT 2

US-09-252-991A-10289  
; Sequence 10289, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10289  
; LENGTH: 840  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10289

Query Match 2.8%; Score 41.2; DB 4; Length 840;  
Best Local Similarity 50.5%; Pred. No. 0.019;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 572 TCACGGCGAGGTCTCTCCGCGAGGCGTGAATCTCATGCGCCACGAGGCTTCGCGGATCA 631  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTGCGGTCCAGT 248  
Db 632 TCTGCTGACGCCCGGCCGGAATCATGTGACGCGAGCACTGTGCTGCGGGCATCGG 691  
QY 249 AGTTGCCGTGGAACCGACCCCTTTGTGTCCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 692 AGAGGATCTTATGAAGCTTTCGCTCTGCTGTTGATCTTCGCCGGCTGTGCGGCTGA 751  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 752 ACGGGAAGCGTCCGACCT 769

## RESULT 3

US-09-252-991A-10493  
; Sequence 10493, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10493  
; LENGTH: 1260  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10493

Query Match 2.8%; Score 41.2; DB 4; Length 1260;  
Best Local Similarity 50.5%; Pred. No. 0.024; 98; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 114 TCACGGCGAGGTCTCTCCGCGAGGCGTGAATCTCATGCGCCACGAGGCTTCGCCGATCA 173  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTTGGGTCCAGT 248  
Db 174 TCTGCTGACGCCCGGCCGATCATGTGACGCGGAGCACCCTGTGCTGCGGGCATCGG 233  
QY 249 AGTTGCCGTGGAACCGACCCCTTTGTGTCCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 234 AGAGGATCTTATGAAGCTTTCGCTCTGCTGTTGATCTTCGCCGGCTGTGCGGCTGA 293  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 294 ACGGGAAGCGTCCGACCT 311

## RESULT 4

US-09-252-991A-10567/c  
; Sequence 10567, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10567  
; LENGTH: 1635  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10567

Query Match 2.8%; Score 41.2; DB 4; Length 1635;  
Best Local Similarity 50.5%; Pred. No. 0.027; 98; Indels 0; Gaps 0;  
Matches 100; Conservative 0; Mismatches 98; Indels 0; Gaps 0;  
QY 129 TCGGGTCGATGATCTCCGCGAAGCGTCCGGTGTCTGACCCAGGAGCTGTGTGCGGTC 188  
Db 1307 TCACGGCGAGGTCTCTCCGCGAGGCGCTGAATCTCATGCGCCACGAGGCTTCGCCGATCA 1248  
QY 189 TCTGACCGCTCCCTCGTCGATCGCTGCTACTGTAAGCGCCCGACGCTTGGGTCCAGT 248  
Db 1247 TCTGCTGACGCCCGGCCGATCATGTGACGCGGAGCACCCTGTGCTGCGGGCATCGG 1188  
QY 249 AGTTGCCGTGGAACCGACCCCTTTGTGTCCTTACCTGCGCGCCACCTGGAGGAATG 308  
Db 1187 AGAGGATCTTATGAAGCCTTTCGCTCTGCTGTTGATCTTCGCCGGCTGTGCGGCTGA 1128  
QY 309 TCGGAACGCTTACGGCAT 326  
Db 1127 ACGGGAAGCGTCCGACCT 1110

## RESULT 5

US-09-252-991A-10674/c  
; Sequence 10674, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136



; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15895  
; LENGTH: 1056  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15895

Query Match 2.6%; Score 38; DB 4; Length 1056;  
Best Local Similarity 48.6%; Pred. No. 0.2;  
Matches 104; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 145 CGCGAACGGTGGGGTCTCGACCCAGGAGTGGTGGGGTCTCTGACCGCTGCCCTC 204  
Db 592 CACAGGCGGTGGGGTACGCGCAAGGGTGTTCGAGGCGCTCGGCACCGTTTC 651  
QY 205 GTGATCGGCTGCATCTGAAGCCCGACGCTGGGTTCACGTAGTTGCCGTGGAAC 264  
Db 652 AACGTACCGCTGCAAGATGCGCGCCACCTGTCTCAGCGTCAAGAGGTTTGGTGGCATC 711  
QY 265 GACCCCTTTGTGTCGCTTACTTGGCGCCACCTTGGAGGAATGTCGGAACGCTTACGC 324  
Db 712 GATGCTTCGGCGCGCTGCCCTGAGGCTCGGTGGCGGAGGCGGCGAGGTATCGG 771  
QY 325 ATCTCTACGACCTAGTCGAGGGCGACTATTTC 358  
Db 772 ATCGCGCGGACGGGTGCGCGGCTCGAACGGC 805

## RESULT 9

US-09-252-991A-15868  
; Sequence 15868, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15868  
; LENGTH: 3357  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15868

Query Match 2.6%; Score 38; DB 4; Length 3357;  
Best Local Similarity 48.6%; Pred. No. 0.36; Indels 0; Gaps 0;  
Matches 104; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 145 CGCGAACGGTGGGGTCTCGACCCAGGAGTGGTGGGGTCTCTGACCGCTGCCCTC 204  
Db 287 CACAGGCGGTGGGGTACGCGCAAGGGTGTTCGAGGCGCTCGGCACCGTTTC 346  
QY 205 GTGATCGGCTGCATCTGAAGCCCGACGCTGGGTTCACGTAGTTGCCGTGGAAC 264  
Db 347 AACGTACCGCTGCAAGATGCGCGCCACCTGTCTCAGCGTCAAGAGGTTTGGTGGCATC 406  
QY 265 GACCCCTTTGTGTCGCTTACTTGGCGCCACCTTGGAGGAATGTCGGAACGCTTACGC 324  
Db 407 GATGCTTCGGCGCGCTGCCCTGAGGCTCGGTGGCGGAGGCGGCGAGGTATCGG 466  
QY 325 ATCTCTACGACCTAGTCGAGGGCGACTATTTC 358  
Db 467 ATCGCGCGGACGGGTGCGCGGCTCGAACGGC 500

## RESULT 10

US-09-252-991A-4342/c  
; Sequence 4342, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4342  
; LENGTH: 1257  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4342

Query Match 2.6%; Score 37.8; DB 4; Length 1257;  
Best Local Similarity 50.6%; Pred. No. 0.25;  
Matches 117; Conservative 0; Mismatches 112; Indels 2; Gaps 1;  
QY 149 GAACGGTGGGGTCTCGACCCAGGAGTGGTGGGGTCTCTGACCGTCCCTCGTCG 208  
Db 534 GAGCTGATGACATCTGCAAGAGGAGGAGGCGGCGGCGCAAGGTCTCTGGCCTACACG 475  
QY 209 ATCGGCTGCATCTGAACCGCCGACCTTGGGTCCACGTAGTTGCGTGGAAACCGACC 268  
Db 474 GTCTATACCGCGCGCGCACACCTCGGCTGAGGTGTTGCT--TGGAGCAGGAAG 417  
QY 269 CTTTGTGTCGCTTACTTGGCGCCACCTTGGAGGAATGTCGGAACGCTTACGSCATCT 328  
Db 416 GCTTCAAGTGGCGGTCTCGCGCGAGCGTGGATGCCAGCGCGCGGAAGACTGGATCG 357  
QY 329 CCTACGACCTAGTCGAGGGCGACTATTGTTTAAACGAAGGGGCGCAAGCTGG 379  
Db 356 CCGAGCAACTGGACCGTGGCATCGACGTCTCATCAACACCGCGAGTTGG 306

## RESULT 11

US-09-252-991A-4513  
; Sequence 4513, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 4513  
; LENGTH: 1353  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-4513

Query Match 2.6%; Score 37.8; DB 4; Length 1353;  
Best Local Similarity 50.6%; Pred. No. 0.26;  
Matches 117; Conservative 0; Mismatches 112; Indels 2; Gaps 1;  
QY 149 GAACGGTGGGGTCTCGACCCAGGAGTGGTGGGGTCTCTGACCGTCCCTCGTCG 208  
Db 823 GAGCTGATGACATCTGCAAGAGGAGGAGGCGGCGGCGCAAGGTCTCTGGCCTACACG 882  
QY 209 ATCGGCTGCATCTGAACCGCCGACGTTGCGGTCCACGTAGTTGCGTGGAAACCGACC 268

Db 883 GTCTATACCGGACGCGGACACACGTCGCGCTGAAGGTGTC--TGAGCAGGAAG 940  
QY 269 CTTTGTGTCGCTTACCTGCGCGCCACCTGAGGAATGTCGGAACGCTTACGGCACT 328  
Db 941 GCTTCAAGTGGCGGTGCTGCGCGAGCGTGGATGCCAGCGCGCGGAAGACTGGATCG 1000  
QY 329 CCTACGACTAGTCAGGCGGACGATTTGCTTAACAAGGGGCCAAGCTGG 379  
Db 1001 CCGAGCAATGGACCGTGCAATCGAGCTGCTCATCACCAACCCGAGTTGG 1051

RESULT 12  
US-09-252-991A-15497/c  
; Sequence 15497, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15497  
; LENGTH: 1419  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15497

Query Match 2.6%; Score 37.8; DB 4; Length 1419;  
Best Local Similarity 50.3%; Pred. No. 0.27;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGTCCTCCGCGAACGTCGCGGTGCTGCACTGAAACCGCCCGGCTTGCCTCC 184  
Db 924 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGCGCTGCGCCACGGAAGTCCAGGCC 865

QY 185 GGTCTCTGACCGCTGCGCTGTCGATCGGTGCTGCACTGAAACCGCCCGGCTTGCCTCC 244  
Db 864 AGGAGAGCATGGAAACTCGCGCTCTGATGTCACGCGGTATCGCGTTCGCGTAG 805

QY 245 ACGTAGTTCGCGTGGAAACCGACCCCTTTGTCGCTTACCTGCGCGCCACCTCGGAGG 304  
Db 804 ACGCCGTTGAAGCGCACGCGGAGGCGTGTCTCGCGAAGCGCGCGCGATGTCGAGG 745

QY 305 AATGT 309  
Db 744 TGAGT 740

RESULT 13  
US-09-252-991A-15291  
; Sequence 15291, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15291  
; LENGTH: 2016

; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15291

Query Match 2.6%; Score 37.8; DB 4; Length 2016;  
Best Local Similarity 50.3%; Pred. No. 0.32;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGTCCTCCGCGAACGTCGCGGTGCTGCACTGAAACCGCCCGGCTTGCCTCC 184  
Db 1520 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGCGCTGCGCCACGGAAGTCCAGGCC 1579

QY 185 GGTCTCTGACCGCTGCGCTGTCGATCGGTGCTGCACTGAAACCGCCCGGCTTGCCTCC 244  
Db 1580 AGGAGAGCATGGAAACTCGCGGTCTGATGTCGACGCGGTATCGCGTTCGCGTAG 1639

QY 245 ACGTAGTTCGCGTGGAAACCGACCCCTTTGTCGCTTACCTGCGCGCCACCTCGGAGG 304  
Db 1640 ACGCGTTGAAGCGCACGCGCGGAGGCTTGTCTCGCGAAGCGCGCGCGATGTCGAGG 1699

QY 305 AATGT 309  
Db 1700 TGAGT 1704

RESULT 14  
US-09-252-991A-15550/c  
; Sequence 15550, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 15550  
; LENGTH: 2508  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-15550

Query Match 2.6%; Score 37.8; DB 4; Length 2508;  
Best Local Similarity 50.3%; Pred. No. 0.35;  
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

QY 125 TGCCTCGGTGATGTCCTCCGCGAACGTCGCGGTGCTGCACTGAAACCGGAGCTTGGTGTCC 184  
Db 1154 TCCTTCTGTACAGCAGGTGCTGGACAGCGCGCGCTGCGCCACGGAAGTCCAGGCC 1095

QY 185 GGTCTCTGACCGCTGCGCTGTCGATCGGTGCTGCACTGAAACCGCCCGGAGCTTGCCTCC 244  
Db 1094 AGGAGAGCATGGAAACTCGCGCTCTGATGTCGACGCGGTATCGCGTTCGCGTAG 1035

QY 245 ACGTAGTTCGCGTGGAAACCGACCCCTTTGTCGCTTACCTGCGCGCCACCTCGGAGG 304  
Db 1034 ACGCGTTGAAGCGCACGCGGAGGCTTGTCTCGCGAAGCGCGCGCGATGTCGAGG 975

QY 305 AATGT 309  
Db 974 TGAGT 970

RESULT 15  
US-09-252-991A-4633/c  
; Sequence 4633, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2004, 03:13:49 ; Search time 757.822 Seconds  
(without alignments)  
9773.579 Million cell updates/sec

Title: US-10-668-047-1

Perfect score: 1461

Sequence: 1 gtgcatcgatcgccagcagc.....ttccggaatcgctcgta 1461

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT NEW PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06 NEW PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06 PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07 NEW PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08 NEW PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08 PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US09 NEW PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10A PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10B PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10C PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10 NEW PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60 NEW PUB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US60 PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.8	3.1	536	17	US-10-338-110-119
2	44.2	3.0	927	17	US-10-437-963-31196
3	44	3.0	3316	16	US-10-369-493-41493
4	43.8	3.0	1966	16	US-10-369-493-31580
5	42.8	2.9	2889	17	US-10-437-963-55282
6	42.2	2.9	1947	16	US-10-381-779-10
7	41.2	2.8	780	15	US-10-156-761-6003
8	41.2	2.8	1425	13	US-10-282-122A-14027
9	41.2	2.8	9025608	15	US-10-156-761-1
10	41	2.8	1311	13	US-10-282-122A-14145
11	40.6	2.8	2871	16	US-10-369-493-47097
12	39.8	2.7	1401	15	US-10-156-761-6097
13	39.8	2.7	9025608	15	US-10-156-761-1
14	39.6	2.7	1503	13	US-10-282-122A-13659

C 15	39.4	2.7	987	13	US-10-282-122A-31696	Sequence 31596, A
C 16	39.4	2.7	1119	13	US-10-282-122A-11706	Sequence 11706, A
C 17	39	2.7	975	17	US-10-437-963-52538	Sequence 52538, A
C 18	39	2.7	1287	15	US-10-156-761-6510	Sequence 6510, Ap
C 19	39	2.7	13494	15	US-10-156-761-2355	Sequence 2355, Ap
C 20	38.8	2.7	2541	15	US-10-156-761-2855	Sequence 2855, Ap
C 21	38.8	2.7	9810	10	US-09-373-658-35	Sequence 35, Appl
C 22	38.8	2.7	9810	11	US-09-989-687-35	Sequence 35, Appl
C 23	38.2	2.6	597	17	US-10-437-963-18193	Sequence 18193, A
C 24	38	2.6	1051	17	US-10-767-701-13578	Sequence 13578, A
C 25	38	2.6	1135	13	US-10-425-114-20794	Sequence 20794, A
C 26	38	2.6	1584	16	US-10-369-493-42282	Sequence 42282, A
C 27	37.6	2.6	1044	15	US-10-156-761-2398	Sequence 2398, Ap
C 28	37.6	2.6	1272	15	US-10-369-493-44349	Sequence 44349, A
C 29	37.6	2.6	1338	15	US-10-156-761-2638	Sequence 2638, Ap
C 30	37.4	2.6	1332	17	US-10-437-963-58487	Sequence 58487, A
C 31	37.2	2.5	549	17	US-10-437-963-57513	Sequence 57513, A
C 32	37.2	2.5	687	13	US-10-282-122A-26299	Sequence 26299, A
C 33	37.2	2.5	690	13	US-10-282-122A-28509	Sequence 28509, A
C 34	37.2	2.5	1341	15	US-10-156-761-1069	Sequence 1069, Ap
C 35	37	2.5	1188	15	US-10-156-761-5738	Sequence 5738, Ap
C 36	37	2.5	1316	17	US-10-437-963-98904	Sequence 98904, A
C 37	36.8	2.5	2029	17	US-10-437-963-91052	Sequence 91052, A
C 38	36.6	2.5	1049	13	US-10-142-426-358	Sequence 358, App
C 39	36.6	2.5	1049	15	US-10-123-155-358	Sequence 358, App
C 40	36.6	2.5	1049	15	US-10-146-731-358	Sequence 358, App
C 41	36.6	2.5	1049	15	US-10-140-472-358	Sequence 358, App
C 42	36.6	2.5	1049	15	US-10-141-761-358	Sequence 358, App
C 43	36.6	2.5	1049	15	US-10-142-885-358	Sequence 358, App
C 44	36.6	2.5	1049	15	US-10-158-790-358	Sequence 358, App
C 45	36.6	2.5	1049	16	US-10-137-871-358	Sequence 358, App

#### ALIGNMENTS

#### RESULT 1

US-10-338-110-119  
; Sequence 119, Application US/10338110  
; Publication No. US20040023254A1  
; GENERAL INFORMATION:  
; APPLICANT: Fuhrmann, Jeffrey J.  
; APPLICANT: Romesser, James A.  
; TITLE OF INVENTION: A Method To Assess Quorum Sensing Potential Of Microbial  
; TITLE OF INVENTION: Communities  
; FILE REFERENCE: HER-0056  
; CURRENT APPLICATION NUMBER: US/10/338,110  
; CURRENT FILING DATE: 2003-01-07  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 119  
; LENGTH: 536  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus Sequence  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (7)..(7)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (21)..(21)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (69)..(69)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:

```
NAME/KEY: misc feature
LOCATION: (87)..(87)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (213)..(213)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (226)..(226)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (261)..(261)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (273)..(273)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (347)..(347)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (357)..(357)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (398)..(398)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (419)..(419)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (509)..(509)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (517)..(517)
OTHER INFORMATION: n is a, c, g, or t
US-10-338-110-119
```

```
Query Match 3.1%; Score 44.8; DB 17; Length 536;
Best Local Similarity 14.0%; Pred. No. 0.00051;
Matches 43; Conservative 151; Mismatches 114; Indels 0; Gaps 0;

QY 1 GTGCATCCGATCCGACCTGAACTCGCGCAAGCTGCTCTCGCAACTGGACCCC 60
DB 228 SBKCRVSRVVRVMYRCSSMVSATCCVMRSNTTGGARHTKWCNGMTWCRSICMR 287

QY 61 ACTACTCAAGCGTGCTAGGCGAGTTCTTCACTCCCATGAAGCGCGCCAGCTGATGGCT 120
DB 288 CRHSCSSAMYKGCSTRTGMVKWRTYTSKSDVMRBCSTSVARWSMGSDSSYVWSBGGN 347

QY 121 TCAATGCTTCGGTGTGATCTCCGCGGAACGGTGGCGGTGCTCGACCCAGAGCTGGT 180
DB 348 GCVRKEVSNSTRTSGSVRSYRSHCRBSRSATGGWCGSYWKVTCVDNMSAWCGGB 407

QY 181 GTCGGCTCTGACCGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
DB 408 RTWDSBHSANCGSVYGGYCGSYHSHVHVRBRVRDMHVRTSSTSGCSYIG 467

QY 241 GTCCAGCTAGTTCGCGTGAACCGACCCCTTTGCTGCTTACCTGCGCGCCACCTG 300
DB 468 GAYCRVBADCRMDSCVTVRNSVHBYKYWSCKYRSMHYCNRRHBBGVBNVYKVYGV 527

QY 301 GAGGAATG 308
DB 528 DRBYKVTG 535
```

```
RESULT 2
US-10-437-963-31196/c
Sequence 31196, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 31196
LENGTH: 927
TYPE: DNA
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_35525C.1
US-10-437-963-31196

Query Match 3.0%; Score 44.2; DB 17; Length 927;
Best Local Similarity 54.7%; Pred. No. 0.0011;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 136 GATGATCTCCGCGAAGCGTGGCGGTGCTCGACCCAGGAGCTGGTGTCCGGTCTCTGACC 195
DB 834 GAGGACCGCGCGCTTCAGGAGGGTGTGAAGACCCAGCGCGCATGCCGCTGCACCCGCC 775

QY 196 GTGCGCCTCGTCGATCGGCTGCATCTGAACGCCCGCGAGTTCGGTCCACGTAGTTGCC 255
DB 774 GCCGCGACGCGCGCGAGTGCATCCACACGCGCGCGCGTGGCAGAGAGGTTGCACCC 715

QY 256 GTGGAACCGACCCCTTTCTGTCGCTTACCTGCGGCGCCAC 296
DB 714 GAGGACCGCGCACTGCGACGCGCGCGCGCGCGCGCGCCAC 674

RESULT 3
US-10-369-493-41493/c
Sequence 41493, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 41493
LENGTH: 3316
TYPE: DNA
ORGANISM: SPHINGOMONAS
US-10-369-493-41493

Query Match 3.0%; Score 44; DB 16; Length 3316;
Best Local Similarity 47.2%; Pred. No. 0.0027;
Matches 134; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
```

[illegible]

```

RESULT 4
US-10-369-493-31580
; Sequence 31580, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31580
; LENGTH: 1966
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31580

```

Query Match	3.0%;	Score 43.8;	DB 16;	Length 1966;
Best Local Similarity	49.0%;	Pred. No. 0.0023;		
Matches 117;	Conservative 0;	Mismatches 122;	Indels 0;	Gaps 0;
QY	142	CTCCGCGGAACGGTGC	CGGGTGTCTGACCCAGGAGCTGGTTCGGGCTCTCTGACCGCTGCC	201
Db	1313	CGCTTCGTGATCGACCGGGCGGGGCTCGTGGGGCCGATGGCGGACCCATGTCGGGGGCC		
QY	202	CTCGTGTATCGGCTGCATCTGAACCGCCCGAGCTTGCGGTCTCCACGTPAGTTGCGGTGGAA		261
Db	1373	TTCCAGGTTTGGCTTCATCTTCGCTGCCCAATGACCCGTGATGGCGCGGCCGACGAG		1432
QY	262	ACCGACCCCTTTGTCTGTCGCTTACCTCGCGGCCACCCCTCGAGAGAA	TGTCGGAACGCTTAC	321
Db	1433	CGCGAGCTCATCCATGATCGCCACCGCGTGGCGCTTCGACGAGGGCCCATCGCCTTC		1492
QY	322	GGCATCTCTACACCTAGTCGAGGGGAGCTATTTCGTTAA	CCAGGGGCCAAGCTGGA	380
Db	1493	CGCTTCCCGCGGGCGAGGGGTGGCGTTCGAGATGCCGAGGCGCGGAGACGTCGCTGGA		1551

RESULT 5  
US-10-437-963-55282  
; Sequence 55282, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.

```

: APPLICANT: Zhou, Yihua
: APPLICANT: Cao, Yongwei
: APPLICANT: Wu, Wei
: APPLICANT: Boukharov, Andrey A.
: APPLICANT: Barbazov, Brad
: APPLICANT: Li, Ping
: TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53221)B
: CURRENT APPLICATION NUMBER: US/10/437,963
: CURRENT FILING DATE: 2003-05-14
: NUMBER OF SEQ ID NOS: 204966
: SEQ ID NO 55282
: LENGTH: 2889
: TYPE: DNA
: ORGANISM: Oryza sativa
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT4530_57303C.1
US-10-437-963-55282

```

Query Match	2.9%;	Score 42.8;	DB 17;	Length 2889;
Best Local Similarity	46.1%;	Pred. No. 0.0064;		
Matches 143;	Conservative 0;	Mismatches 167;	Indels 0;	Gaps 0;

  

QY	34	CAAGCTGCTCTCGGCAAACTGGACCCACACTACTCAAGCGGTGCTAGGCGAGTCTTCTACT	93
Db	286	CAGAGATTGCCACCTCTCGACGGCGCTCTCTCTCGCGTGGAGCGGTTCCTGCTCGGC	345
QY	94	CCCATGAAGCCCGCCACGCTGATGGCTTCAATGCTTCGGGTTCGATGATCTCCGCGGAACG	153
Db	346	CCGCCGGGTGCTCAGCAGATCTCCCTGGCGCTCGCGGCTGCGGCTCCGCGCCGAG	405
QY	154	GTCCGGGTGCTGCACCCAGAGCTGTGTTCGGGTCTCTGACCGCTGCCCTGTCGATCGG	213
Db	406	GGCGGGGTGACGGGCTCTTCGCGGATCGGCACCTCCCGACCCGCGCTGATGGAG	465
QY	214	CTGCATACTGAACGCCCGGAGGTTCGGTTCACAGTAGTTGCCGTGGAAACGACCCCTTT	273
Db	466	CTGCTTACTGTGCTCCCGAGGAGGTGTCGAGGACCAAGACGCGGGGACACCGGGGTGGAC	525
QY	274	GTGCTGCTTTACCTGTGCGCGCACCTCTGGAGGAATGTTCGGAACGTTTACGGCATCTCCTAC	333
Db	526	GCCGCGGCGGTGTCAGGTTTCAACCGCGAGCTCATGGCGCACGGCCCGGCTCGAG	585
QY	334	GACCTAGTCG	343
Db	586	TTCTCTCTCG	595

```

RESULT 6
US-10-381-779-10
; Sequence 10, Application US/10381779
; Publication No. US200302197981
; GENERAL INFORMATION:
; APPLICANT: Gokarn, Ravi R
; APPLICANT: Jessen, Holly
; APPLICANT: Zidwick, Mary Jo
; TITLE OF INVENTION: Isoprenoid Production
; FILE REFERENCE: 12904/002US1
; CURRENT APPLICATION NUMBER: US/10/381,779
; CURRENT FILING DATE: 2003-03-28
; PRIOR APPLICATION NUMBER: PCT/US01/30328
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: 60/236,580
; PRIOR FILING DATE: 2000-09-29
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-381-779-10

```

Query Match 2.8%; Score 42.2; DB 16; Length 1947;  
 Best Local Similarity 48.5%; Pred. No. 0.0082;  
 Matches 116; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 142 CTCGCCGGAACGGTCCGGGTGCTCGACCCAGGAGTGGTGTGGGTCTCTGACCGCTGCC 201  
 DB 1255 CGCTTCGTGATGACCGGGGGGGTCTGTGGGGCCGATGGCGACCCATCGGGGGCC 1314

QY 202 CTCGTCGATCGCTCATCTGAAACGCGCCGAGCTGGTGTGCGGTCCAGTAGTTCCTGGAA 261  
 DB 1315 TTGACGCTTGGTTCATCATCTTCGTCGCCCAACATGACCGTGTATGGCGCGCCGACGAG 1374

QY 262 ACCGACCCCTTGTGCTGCTTACCTGGCGCCACCTCGAGGAATGTGGAACGCTTAC 321  
 DB 1375 GCCGAGCTCATCCATGATATGCCACCGCGGTGGCTTCGGGAGGGCCCATCGCTTC 1434

QY 322 GGCATCTCTACGACTAGTCTGAGGGCGACTATTTGCTTAAACAAAGGGCCAAAGTGA 380  
 DB 1435 CGCTTCGCCGGGGGAGGGGTGGGGTGGAGTCCGAGTCCCGAGCGCGGACGGTGTGGA 1493

RESULT 7  
 US-10-156-761-6003  
 ; Sequence 6003, Application US/10156761  
 ; Publication No. US20030119018A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OMURA, SATOSHI  
 ; APPLICANT: IKEDA, HARUO  
 ; APPLICANT: ISHIKAWA, JUN  
 ; APPLICANT: HORIKAWA, HIROSHI  
 ; APPLICANT: SHIBA, TADAYOSHI  
 ; APPLICANT: SAKAKI, YOSHIYUKI  
 ; APPLICANT: HATTORI, MASAHIRA  
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 ; FILE REFERENCE: 249-262  
 ; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204089  
 ; PRIOR FILING DATE: 2001-05-30  
 ; PRIOR APPLICATION NUMBER: JP 2001-272697  
 ; PRIOR FILING DATE: 2001-08-02  
 ; NUMBER OF SEQ ID NOS: 15109  
 ; SEQ ID NO 6003  
 ; LENGTH: 780  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces avermitilis  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(780)  
 US-10-156-761-6003

Query Match 2.8%; Score 41.2; DB 15; Length 780;  
 Best Local Similarity 47.3%; Pred. No. 0.011;  
 Matches 124; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

QY 130 CGGGTGTGATCTTCGCGGAACGGTGGGGTGTGACCCAGGAGTGTGTGCGGTCT 189  
 DB 85 CGTGGCGCGCGCGCGGGGATGTTGTGCGCGCGGTTCGCGGACTGTGTGCGGGCTG 144

QY 190 CTGACCGCTGCCCTCGTTCGATCGGCTGCATCTGAAACGCGCCCGACGTTGCGGTCCAGTA 249  
 DB 145 CTGCGCGCGCGCTCGGACGCTGCTCTCTGCTGGCGCGAGCGCGTGTGCTCGCGCC 204

QY 250 GTTGGCGTGAACCGACCCCTTTGTGCTGCTTACCTGGCGGCCACCTGGAGGATGT 309  
 DB 205 GTGGCCACCGCGCGTGGCGCGCTCTCACCGCGGGTCCACTCTGACGGGCTGCC 264

QY 310 CGGACGCTTACGCACTCTCTACGACTAGTCTGAGGGGCACTATTTGCTTAAACAGGG 369  
 DB 265 GACACCGGAGCGCTTCGGGAGCGGAAGCCCGCGAGGACGGCTGCGGATCATGAG 324

QY 370 GCCAAGCTGGATGCCCGCTTCG 391  
 |||||

Db 325 CAGTCGACATCGGCCCGTTCG 346

RESULT 8  
 US-10-282-122A-14027/C  
 ; Sequence 14027, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/230,335  
 ; PRIOR FILING DATE: 2000-09-06  
 ; PRIOR APPLICATION NUMBER: 60/230,347  
 ; PRIOR FILING DATE: 2000-09-09  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/267,636  
 ; PRIOR FILING DATE: 2001-02-09  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 14027  
 ; LENGTH: 1425  
 ; TYPE: DNA  
 ; ORGANISM: Burkholderia mallei  
 US-10-282-122A-14027

Query Match 2.8%; Score 41.2; DB 13; Length 1425;  
 Best Local Similarity 52.3%; Pred. No. 0.015;  
 Matches 91; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 125 TGCTTCGGTTCGATATCTCCGGGAACGGTGGGGTGTGACCCAGGAGCTGGTGTG 184  
 DB 236 TGGCGCGCTCGATACGTCGCTGTGCTGTGTCGACGTCGACGAAGCGGGCGG 177

QY 185 GGTCTCTGACCGCTGCCCTCGTGCATCGGCTGCATACCTGAACCGCCCGACGTTGCGGTCC 244  
 DB 176 AGCGCGCGCGCTTCGCTCGACACCTGCTCGGCAAGCCAGTTCGTTCAGATTGCCGATC 117

QY 245 ACCTAGTTCGCTGGAACCGACCCCTTTGCTCGTCTTACCTGCGCGCCACCC 298  
 DB 116 GCGACGATCGGTGCGCGGTATCGGACGTCGCGCTCGACGCTCGCGCGGCC 63

RESULT 9  
 US-10-156-761-1/c  
 ; Sequence 1, Application US/10156761

```
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      2.8%; Score 41.2; DB 15; Length 9025608;
Best Local Similarity 47.3%; Pred. No. 2.3; Indels 0; Gaps 0;
Matches 124; Conservative 0; Mismatches 138;

Qy 130 CGGTGCGATGATCTCCGCGAACGCTGCGGTGCTCGACCCAGGAGCTGTGTCGGGTCT 189
Db 7256910 CGTCCGCGCGCGCGCGGAGATGTGTGCGCGCGGTGCGGACTGTGTCGGGTG 7256851

Qy 190 CTGACCGCTGCTGCTGCTGATCGGTGCTGATGAAAGCCCGGACGTTGCGGTCCACGTA 249
Db 7256850 CTCGCCGCGGCGCTCGGCACGCTGCTGCTGCTGCGCGGAGCCCGTTGCTCGCGCC 7256791

Qy 250 GTTCCGCTGGAACCGACCCCTTTGCTGCTTACCTGCGCGGACCACTTGTCTTAACCAAGG 309
Db 7256790 GTGGCCACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 7256731

Qy 310 CGGAACGCTTACGGCATCTCTTACGACCTAGTCTGAGGCGGCACTATTGCTTAACCAAGG 369
Db 7256730 GACACCGGAGCGGCTCGGCGAGCGCAAGCCCGCGAGGACGCGTGGCGATCATGAG 7256671

Qy 370 GCCAGCTGATGCGCGCTTCG 391
Db 7256670 CAGTCGACATCGGCGCGCTTCG 7256649

RESULT 10
US-10-282-122A-14145
Sequence 14145, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
```

```
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 14145
LENGTH: 1311
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-14145

Query Match      2.8%; Score 41; DB 13; Length 1311;
Best Local Similarity 46.9%; Pred. No. 0.017; Indels 0; Gaps 0;
Matches 128; Conservative 0; Mismatches 145;

Qy 108 CACGCTGATGCTTCAATGCTTCGGGTGCGATGATCTCCGCGAACGCTGCGGTGCTCGA 167
Db 696 CACGCGCTCGCGCGGATGCGGTGCGCGGAGATGATGACGACGCGCGGCTCGCGA 755

Qy 168 CCCAGGAGCTGCTGCGGTCTCTGACCGCTGCCCTGTCGATCGGTGTCATGATGAACG 227
Db 756 CGGCGCTGTCGCGACGCTGATTCGATGACGACATGTCGACCATGCTCGTGTTCG 815

Qy 228 CCCGAGCTTGGCTCCAGCTAGTTCGCTGGAACGACCCCTTTGTCGTCCTTACCT 287
Db 816 GCACGGGAGCGCCACGCGAGGCGCGCTGCGCGTGCACCGGTGCGAGCGGATCGC 875

Qy 288 GCGCGCACCTGAGGAATGTGGAACGCTTACGGCATCTCTTACGACCTAGTTCGAGGG 347
Db 876 GCGCACTTACGCGCGGTGTCCCGAACGCGCCACCGTGCACCGAGCTCGCGCGCG 935

Qy 348 CGACTATTGCTTAACCAAGGGGCAAGCTGGA 380
Db 936 GCGGGGCTTTGCTGCTGCGCGCGAGCTCGA 968

RESULT 11
US-10-369-493-47097/c
Sequence 47097, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
```

```
; SEQ ID NO 47097
; LENGTH: 2871
; TYPE: DNA
; ORGANISM: Deinococcus radiodurans
US-10-369-493-47097

Query Match      2.8%; Score 40.6; DB 16; Length 2871;
Best Local Similarity 53.5%; Pred. No. 0.036; 74; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 0;

QY 131 GGGTCGATGATCTCCGCGGAACGCGTGGCGGTGCTCGACCCAGAGCTGGTGTGCGGTCTC 190
    |||||
DB 608 GCGGGTTGAGTTTTCGATCAGCGCGCGCTCATCTGCTCAGGCGGTAGCGGCC 549
    |||||

QY 191 TGACCGTGCCTCGTGTGATCGGTGCTGATACTGAACGCCGCCGACGTTGGGTCCAGTAG 250
    |||||
DB 548 TGACCGCGCCCTTCTCGATGGCTCTGAAACCGCGCGTTCGTTGGCGCGCAGGTAG 489
    |||||

QY 251 TTGCGGTGGAACCGACCCCTTTGTCGTGCTTACTGC 289
    |||||
DB 488 TTGTACTGAAACCGATGGCGCGCAGTAGGTGCTTGC 450
    |||||

RESULT 12
US-10-156-761-6097
; Sequence 6097, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6097
; LENGTH: 1401
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1401)
US-10-156-761-6097

Query Match      2.7%; Score 39.8; DB 15; Length 1401;
Best Local Similarity 44.7%; Pred. No. 0.045;
Matches 155; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 127 CTTGCGGTGATGATCTCCGCGGAACGCGTGGCGGTGCTCGACCCAGAGCTGGTGTGCGG 186
    |||||
DB 883 CTTGCGGACGCGGCCCTTGGCGGTGCTCGACCCAGAGCTGGTGTGCGG 942
    |||||

QY 187 TCTGTACCGCTGCCCTCGTGTGATCGGTGCTGATCTGAACGCCGCCGACGTTGGGTCCAC 246
    |||||
DB 943 CTGCGGCGCTTCCGCGCGACGCCACCGATCGCGACGCTGCGCGACGTGACGGGTC 1002
    |||||

QY 247 GTAGTTGCCGTGGAACCGACCCCTTTGTGTCCTTACTGCGCGCCACCCCTGGAGAA 306
    |||||
DB 1003 ACCTACATCGAGACTTCAAGGCCACCAACGACGACCGCGGAAGCTCTTTGGCGCG 1062
    |||||

QY 307 TGTGCGAACGCTTACGCGCATCTCTACGACCTAGTCGAGGCGGACTATTTGCTTAACAA 366
    |||||
DB 1063 TACGGTCCATCGTCTGATCGCGGCGGTCTGCCAAGGCGCTCTCTTCGACGAGCTG 1122
    |||||
```

```
QY 367 GGGGCCAAGCTGGATGGCCCGTTCGATCTTGTAAATGCTAATCTCCCTACGAAAGCTT 426
    |||||
DB 1123 GTGGCCAAAGTCGGACACGCGCTCGGGGTGTGGTCTCATCGGTGCGGATCGTGCCTG 1182
    |||||

QY 427 GCTTCAGATTGCTGGCGCGGCTTGCAACGACAGCGCGTGGCGTGA 473
    |||||
DB 1183 ATCCGTGAAGCCCTTTCGCGGACACGCCCGCGGAGGTACCGCTGCTGA 1229
    |||||

RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      2.7%; Score 39.8; DB 15; Length 9025608;
Best Local Similarity 44.7%; Pred. No. 6.8; 192; Indels 0; Gaps 0;
Matches 155; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 127 CTTGCGGTGATGATCTCCGCGGAACGCGTGGCGGTGCTCGACCCAGAGCTGGTGTGCGG 186
    |||||
DB 7361771 CTTGCGGACGCGGCCCTTGGCGGTGCTCGACCCAGAGCTGGTGTGCGG 7361830
    |||||

QY 187 TCTGTACCGCTGCCCTCGTGTGATCGGTGCTGATCTGAACGCCGCCGACGTTGGGTCCAC 246
    |||||
DB 7361831 CTGCGGGCTTTCGCGCGGACGCCACCGCATCGCGACGTCGCCGACGCGGGTC 7361890
    |||||

QY 247 GTAGTTGCCGTGGAACCGACCCCTTTGTGTCCTTACTGCGCGCCACCCCTGGAGAA 306
    |||||
DB 7361891 ACCTACATCGAGACTTCAAGGCCACCAACGACGCGCGAAGCTCTTTGGCGCG 7361950
    |||||

QY 307 TGTGCGAACGCTTACGCGCATCTCTACGACCTAGTCGAGGCGGACTATTTGCTTAACAA 366
    |||||
DB 7361951 TACGGGTCCATCTCTGATCGCGGCGGCTCGCAAGGCGCTCTCTTCGACGAGCTG 7362010
    |||||

QY 367 GGGGCCAAGCTGATGCGCGGTTCGATCTTGTGTAATGCTAATCTCCCTACGAAAGCTT 426
    |||||
DB 7362011 GTGGCCAAAGTCGCGACACAGCGGCTCCGGGTGTGTGTCTCATCGGTGCGGATCGTGCCTG 7362070
    |||||

QY 427 GCTTCAGATTGCTGGCGGCTTGAACGACAGCGGTGCCCGTGA 473
    |||||
DB 7362071 ATCCGTGAAGCCCTTTCGCGGACACGCCCGCGGAGGTACCCGTCGTA 7362117
    |||||

RESULT 14
US-10-282-122A-13659
; Sequence 13659, Application US/10282122A
; Publication No. US20040029129A1
```

GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

FILE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 13659

LENGTH: 1503

TYPE: DNA

ORGANISM: Burkholderia fungorum

US-10-282-122A-13659

Query Match 2.7%; Score 39.6; DB 13; Length 1503;

Best Local Similarity 50.8%; Pred. No. 0.055;

Matches 120; Conservative 0; Mismatches 114; Indels 2; Gaps 1;

QY	138	TGATCTCCGCGAACGGTGGGGTCTCGACCCAGGAGCTGGTGTGGGTCTCTGACCGC	197
Db	431	TCATTGCGGACGCGGTCCACGACTCCACCGGAGCTGCTCACCGGCGCTTCGGCC	490
QY	198	TGCCTCTGTCGATGGGTGTCATCTGAACGCCCGACGTTGCGGTCCACGTAAGTTCGCGT	257
Db	491	TTACCGCGCGCGCGACGGCGTGTGAGCGCGCGAGCTGCTGTATCCGACCTGTCT	550
QY	258	GGAAACCGACCCCTTTCTGTCTTACCTGCGGCGCACCGCTGAGGAATGTCGGAACGC	317
Db	551	GGCTGCGGTGCGCGACGGCGGCAAGCGCGCGGTGAACCGCGCATCACGTCATGA	610
QY	318	TTACGGCATCTCTACGACTAGTCGAGGCGCACTATTTCCTTAACCAAGGCGCA	373
Db	611	CCACGNAACCGTCATGAC--AGTCGACGCGCACACCTGCTGACGACGCGCA	664

RESULT 15

US-10-282-122A-31696/c

; Sequence 31696, Application US/10282122A

; Publication No. US20040029129A1

; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

FILE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 31696

LENGTH: 987

TYPE: DNA

ORGANISM: Pseudomonas putida

US-10-282-122A-31696

Query Match 2.7%; Score 39.4; DB 13; Length 987;

Best Local Similarity 54.5%; Pred. No. 0.051;

Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY	134	TCGATGATCTCCGCGAACGGTGGGGTGTGCGGTGCTCGACCCAGGAGCTGGTGTGGGTCTCTGA	193
Db	584	TTGCTGAATCTCGACGAGCGCGGCGCTGTGTCGACGAGCTGTGTCGCGCAGGGTCTGG	525
QY	194	CCGCTGCCCTTCGTGATCGGCTGCATCTGAACGCCCGGACGTTGCGGTCCACGTAGTTG	253
Db	524	GTGCTGTGTGTCATCCAGTCCAGTCAGCATCTTGGTGTTCGCTGCTGTGACCGCCACTTG	465
QY	254	CCGTGGAACCGACCCCTTTGTCTGT	278
Db	464	CCGTGCGGGCTGACCGCCATGCCTT	440

Search completed: October 2, 2004, 08:33:22

Job time : 780.822 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 16.8208 Seconds  
(without alignments)  
2779.248 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VPIASTETRRQAALGKLDP.....QETLDVAVVRLFCIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	667.5	27.0	507	1 XYORS	site-specific DNA-
2	662.5	26.8	501	1 S35515	site-specific DNA-
3	379	15.4	546	2 AF2268	type II site-speci
4	243.5	9.9	236	2 AE2268	site-specific DNA-
5	184	7.5	421	2 JN0257	site-specific DNA-
6	180.5	7.3	428	2 JH0634	site-specific DNA-
7	169.5	6.9	377	1 S01615	site-specific DNA-
8	165	6.7	487	2 G64577	type I restriction
9	163	6.6	540	1 JU0470	site-specific DNA-
10	162.5	6.6	531	1 XPS7A	site-specific DNA-
11	161	6.5	543	2 H71935	type I restriction
12	151	6.1	561	2 JN0797	site-specific DNA-
13	148	6.0	579	2 I40371	methyltransferase
14	147	6.0	379	1 P64633	site-specific DNA-
15	141	5.7	571	2 H97333	site-specific modi
16	141	5.7	750	2 AH3158	hypothetical prote
17	141	5.7	832	2 F98128	insertion element
18	140	5.7	555	2 E87003	probable DNA methy
19	137	5.5	514	2 F81152	type I restriction
20	135	5.5	553	2 H70977	hypothetical prote
21	134	5.4	514	2 D81952	type I site-specif
22	134	5.4	580	2 JH0224	site-specific meth
23	133.5	5.4	381	2 A71882	type II DNA modifi
24	132	5.3	1181	2 B64516	hypothetical prote
25	131.5	5.3	620	2 JC4925	XmI methyltransfe
26	130	5.3	520	2 S02166	type I site-specif
27	130	5.3	545	2 F64579	site-specific DNA-
28	128	5.2	518	1 F64073	site-specific DNA-
29	126	5.1	545	2 F71932	type II DNA modifi

ALIGNMENTS

RESULT 1

XYOFS

site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) PstI - Providencia ;  
N:Alternate names: modification methylase PstI; restriction-modification system PstI  
C:Species: Providencia stuartii  
C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 05-May-2000  
C:Accession: A00553  
R:Walder, R.Y.; Walder, J.A.; Donelson, J.E.  
J. Biol. Chem. 259, 8015-8026, 1984  
A:Title: The organization and complete nucleotide sequence of the PstI restriction-modif:  
A:Reference number: A92485; MUID:84239756; PMID:6330092  
A:Accession: A00553  
A:Molecule type: DNA  
A:Residues: 1-507 <WAL>  
A:Cross-references: GB:K02081; NID:g150922; PIDN:AAA25672.1; PID:g455323  
A:Experimental source: strain 164  
A:Note: the authors translated the codon CAA for residue 456 as Glu  
C:Comment: Providencia, a genus of enterobacteriacean rods, is sometimes included in the  
C:function: <Met>  
A:Description: This site-specific methylase recognizes the double-stranded hexanucleotid  
y PstI endonuclease (see PIR:NDOFS).  
C:Superfamily: site-specific DNA-methyltransferase (adenine-specific) PstI  
C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 27.0%; Score 667.5; DB 1; Length 507;  
Best Local Similarity 34.2%; Pred. No. 2.2e-43;  
Matches 166; Conservative 87; Mismatches 206; Indels 27; Gaps 10;  
Qy 11 RQALGKLDPTQAVLGQFTTPKKAATLMASMLRVDDLGRVTVRVLDGAGVGSITAAIYD 70  
Db 20 REAANSTLDLTRSKLQGFMSSSAVSELMANLF--ESYVGEHILDRAGVGSUTAAFPV 77

Qy 71 RLHTRPDVAVHVAVTDPFVPLRATLEECR--NAYGISYD--LVEGDYLLNQAG 125

Db 78 NA-TLNGAKSISSTCYEISEVMVNNLLQVLDLCKIRAMEFVNWQKLIIESDFTQASVEQ 136

Qy 126 L-----DGPDLVIANPPYKGLASDSLARLATTARAVDVPNVYVAVFVWRAVISLKEQGR 180

Db 137 LLINYSKPKYKALINPPYLKIAKGRERALLQKVGIEASNLVSFAVALAKQLKSGGEL 196

Qy 181 VFTVPSRWANGPYRQPHWLMTAVSLDILHVPESRKFVFADTKVKQENIVAFSVRPS 240

Db 197 VAITPRSCFNGPYNDFRQKMLDECSLNKIHVNSRKSFAKADNVLGQIYHLTKGETQ 256

Qy 241 SSV--VLSRVAHGEESIASVPSFALVHDEDDDKIVHFA-----ESASVPSAARFLLA 292

Db 257 RKVVTVYSSTCANDINPTIEVFDFEIVKSNNDLFLHIIVTNEQERLANKAGLPCSL 316

Qy 293 DLGIGVSTGKVDPRNRQYLTDLNLDASGVVPMVYQSNIRSGKIDWPQVGARKPGGFVAVE 352

Db 317 DLGIQVSTGKVDPRNRENLISMEY--ISNSVELIFPQHLQRCISIVWPITKAKPNALI-VN 374

probable dimethyla  
probable type II D  
type I site-specif  
type I restriction  
adenine/cytosine D  
vitamin B12-depend  
hypothetical prote  
kasugamycin dimeth  
hypothetical prote  
dimethyladenosine  
hypothetical prote  
hemk protein homol  
protoporphyrinogen  
conserved hypothet  
conserved hypothet  
type I restriction



```
QY 6 STETRRQAALGLDPTTQAVLGGQFTPMKAATLMASMLRVDDLRGTVRVLDPGAGVGSIT 65
Db 8 STDINRINFSSRLNLHRRRELQFLLTAPLARFMVG--QFNSLLGHISLIDPGAGVGSIT 65
QY 66 AALVDRLHATERPDV-AHVHVAVEDTDFVVPVIRATLECRNAY---GI--SYDLVEGDYL 119
Db 66 AAFVERLLANSHEVKSCTFYVEPIFILPKQCLDCCDALENKGAANYCYLCKNF- 124
QY 120 LNOGAKLDGP-----FDLVIANPPYKGLASDSLARLATTARAVDVPNVVAFVVRVAV 171
Db 125 IDANSEINPLSTSAINTFTHAIINPLYKKNKSIEKKIISQMGIEIVNLYSAFVWLTI 184
QY 172 ISLKEQGRGVFIVPRSWANGPPYRQFRHMLMTAVSLD 208
Db 185 LRLVEDGEIVAITPRFCNGSYFRHFRKSFLEQMKLE 221
RESULT 5
JN0257
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) TaqI - Thermus aqua
C:Species: Thermus aquaticus
C>Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 05-May-2000
C:Accession: JN0257; S06298
R:Barany, F.; Slatko, B.; Danzitz, M.; Cowburn, D.; Schildkraut, I.; Wilson, G.G.
Gene 112, 91-95, 1992
A>Title: The corrected nucleotide sequences of the TaqI restriction and modification enz
A:Reference number: JN0257; MUID:92201707; PMID:1551602
A:Accession: JN0257
A:Molecule type: DNA
A:Residues: 1-421 <BAR>
A>Note: this is a revision to the sequence from S06298
R:Slatko, B.E.; Benner, J.S.; Jager-Quinton, T.; Moran, L.S.; Simcox, T.G.; Van Cott, E.
Nucleic Acids Res. 15, 9781-9796, 1987
A>Title: Cloning, sequencing and expression of the Tag I restriction-modification system
A:Reference number: S06298; MUID:98096538; PMID:2827113
A:Accession: S06298
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-12,'A',14-325,'Y',327-351,'GGRSSTSC',360,'RKV' <SLA>
A:Cross-references: GB:Y00499; NID:9288593; PIDN:CAA68551.1; PID:G288594
A>Note: it is uncertain whether Met-1 or Met-31 is the initiator
C:Genetics:
A:Gene: taqIM
C:Function:
A:Description: this enzyme is a methyltransferase of TaqI restriction-modification syste
A>Note: recognition sequence of this enzyme is 5'-TCGA
C:Keywords: methyltransferase; S-adenosylmethionine
Query Match 7.5%; Score 184; DB 2; Length 421;
Best Local Similarity 23.5%; Pred. No. 2.5e-06;
Matches 109; Conservative 58; Mismatches 177; Indels 120; Gaps 20;
QY 26 LGQFFTPMKAATLMASMLRVDDLRGTVRVLDPGAGVGSITAAALVDRLHTRPPVAVHVA 85
Db 18 LGVEVTPPEVDPWVSLAEP--RGG-RVLEPCAHGPFL-----RAFREAHTGYRFGV 69
QY 86 VETDPEW---PYLRATLEECRNAYGISDVLVEGDYLLNQAKLDGPFOLVIANPPYG-- 140
Db 70 VEIDPKALDLPWAEGIL-----ADFLLWEPGE---AFDILLGNPPYGIV 111
QY 141 -----KLASDSLARLATTARAVDVPNVVAFVVRVAVISLKEQGRGVFIVPRSWA 189
Db 112 GEASKYPIHVFAVKDLYKKAFTWKG--KYNLYGAFLEKAVRELLKPGGVLVFVVPVATWL 169
QY 190 NGPYRQFRHMLMTAVSLDILHVFESRTKVFDATKVKQENVIVAFSVRQSSSVLSRV 249
Db 170 VLEDFALLREFLAR-----EGKTSVYVILGEVFPQKVSADVIRFPKSGKGLSLWD 219
QY 250 AHGEESIASVPFSAVHDEDDDKIVHF-AESASVPSAARFTLADLIGIGVSTGKVDREN 308
Db 220 TQESGESGTPILWAEYPHWEGE--IIRFETETRKLEISGMPLGDLFHIRFAARSPFCK 277
QY 309 RQYLTNDLDASGVVPMVYQSNIRSGKIDW-----PQVGARKPQGFVAVEDVALRQL 359
```

```
Db 278 HPAVRKE--PGSLPVLTCRNKLGKWDYKHNKSLGMLWPKERAKELRDFHAT----- 328
QY 360 LPOGSYVVVKQTKAKEDRRRRIAADWAGAS-----RVALDNKTNLYLHESQRP 406
Db 329 ----PHLVV-----AHTKGRVVA--WDERAYPWREEFHLLPKGVRLDPST----- 370
QY 407 LEKNVARGMLMLNSTVLDQY-----FRAFSQHTQVNAAGDLRLP 446
Db 371 -----LVQWLNSEAMQKHVRLTYRDFVPHLTRM--LERLP 404
RESULT 6
JH0634
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) TthHB81 - Thermus a
C:Species: Thermus aquaticus
C>Date: 17-Aug-1992 #sequence_revision 17-Aug-1992 #text_change 05-May-2000
C:Accession: JH0634
R:Barany, F.; Danzitz, M.; Zebala, J.; Mayer, A.
Gene 112, 3-12, 1992
A>Title: Cloning and sequencing of genes encoding the TthHB81 restriction and modificati
A:Reference number: JH0634; MUID:92201699; PMID:1339363
A:Accession: JH0634
A:Molecule type: DNA
A:Residues: 1-428 <BAR>
A>Note: the source was designated as Thermus thermophilus HB8
C:Genetics:
A:Gene: tthHB81M
C:Keywords: methyltransferase; S-adenosylmethionine
Query Match 7.3%; Score 180.5; DB 2; Length 428;
Best Local Similarity 22.6%; Pred. No. 4.7e-06;
Matches 109; Conservative 59; Mismatches 179; Indels 135; Gaps 21;
QY 18 LDPTTQA-VLGGQFTPMKAATLMASMLRVDDLRGTVRVLDPGAGVGSITAAALVDRLHTR 76
Db 7 LPPTASGRSLGRVETP--PGLVRFVGLAEARKGVRLFPACADGPFL-----RAPREA 58
QY 77 PDVAVHVAVEDTDFVY---PYLRATLEECRNAYGISDVLVEGDYLLNQAKLDGPFDLV 133
Db 59 HGTGYRFGVGEIDPHALDLPFWAEG-----VWADFLLWEPGE--AFDLI 100
QY 134 IANPPYKGLASDSLARL-----ATTARAVDVPNVVAFVVRVAVISLKEQGRGVF 182
Db 101 LGNPPYGVGEASKYPIHVLRVKGLYKKTSLTWKGYKYNLYGAFIEKSVRLLEGSTLVF 160
QY 183 IVPRSWANGPYRQFRHMLMTAVSLDILH---VFESRTKVFA-----DTK 224
Db 161 VVPATWLVDLDFSLRSFLAREGRTEVYVILGEVFPGR-KVSAVVLFRKGGKGLALWDTR 219
QY 225 VKQENVI-VAFSVRPOSSSVLSRVANGHEESTASSVPFSAVHDEDDDKIVHFAFSASV 283
Db 220 RDGETFTPLLWSEKPEWKEGIIIRFETGTWREMEASGPFLGSLFH-----I 264
QY 284 PSAARFTLADLIGIGVSTGKVDPRNRYLTNDLDASGVVPMVYQSNIRSGKIDW----- 337
Db 265 PFAARSP-----EFKHPAVQKEPE-PGLVPIVLTGRNLKPGWIDYESNHSG 309
QY 338 ---PQVGARKPQGFVAVEDVALRQLLPQGSYVVKVQTKAKEDRRRVIAAVWDGASVALD 394
Db 310 LWWPKERAKELRDFVATPHL-----VVAHTKGTGK-----VVAW 344
QY 395 NKTNYLHESQRPLEKNVAR-----GLMLNLSNVLDQY-----FRAFSGHTQVNAAGDLRLP 446
Db 345 EKAYPWREEFHLLPKREGVELDPLFLVWLNDSKIOEYVKTLYRDFVPHLTRM--LERIP 402
QY 447 FL 448
Db 403 AL 404
RESULT 7
S01615
```

site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) CviBIII - Chlorelli  
C;Species: Chlorella virus CV-NC1A  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: S01615  
R;Narva, K.E.; Wendell, D.L.; Skrdla, M.P.; van Etten, J.L.  
Nucleic Acids Res. 15, 9807-9823, 1987  
A;Title: Molecular cloning and characterization of the gene encoding the DNA methyltrans  
A;Reference number: S01615; MUID:88096540; PMID:3320956  
A;Accession: S01615  
A;Molecule type: DNA  
A;Residues: 1-377 <NAR>  
A;Cross-references: EMBL:X06618; NID:g60638; PIDN:CAA29835.1; PID:g60639  
C;Genetics:  
C;Superfamily: site-specific DNA-methyltransferase (adenine-specific) CviBIII  
C;Keywords: methyltransferase; restriction modification system; S-adenosylmethionine

Query Match 6.9%; Score 169.5; DB 1; Length 377;  
Best Local Similarity 23.7%; Pred. No. 2.7e-05;  
Matches 75; Conservative 46; Mismatches 123; Indels 73; Gaps 16;

Qy 27 GQFPTPKMAATLWASMLRVDLDLGRVRLDPGAGVGSLLTAALVRLHLTERPDVAHVAV 86  
Db 26 GIFFPTKTVZEKLGFFTEHQNTPGFSILEPSCGTGLIIECVERF---PLASIKGVEL 81  
Qy 87 ETDPFPVVPYLRLATLEECRNAYG-ISYDLVSGDYLLNNOGAKLDGPFDLVIANPPY----G 140  
Db 82 DNDMSTI-----CSKQYAEVNDIVNEDFLLWKG---GKFDFIVGNPPVVRPSG 128  
Qy 141 KLASDSLRLATATARVNDVNVVAVFWRAVIS-LKEQGRGVFIVPRSWANGPYRQFRH 199  
Db 129 YKQDNRLAK-----GRS-----NLVYEFLYKCTEHLKEDGILALIPSTIGNSSFYEP 180  
Qy 200 WLMTAVSILDIHLVPESTRK-VFADTKVKQENVI VAFSVRPSQSSVVLRSVAHGEESIAS 258  
Db 181 LIIT--LDLILS-PEILDKHDFCDTNR-----LCSIVIKNSPGTGKYT-- 220  
Qy 259 SVPSALVHDBDDDKIVHFAESASVPSAARTLADLGIGVSTGKVDPRNQYLTDNLA 318  
Db 221 --YRDVTCRDK-----IPHHGNSYIGSLDLKFKTGFPAWNVAK-FFTCKSE- 263  
Qy 319 SGVVMVYQSNIRSGKI 335  
Db 264 ---IPFTSSNIKLEI 277

RESULT 8  
type I restriction enzyme M protein - Helicobacter pylori (strain 26695)  
C;Species: Helicobacter pylori  
C;Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 17-Mar-2000  
C;Accession: G64577  
R;Tomb, J.F.; White, O.; Kervilave, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, C.; Nature 388, 539-547, 1997  
A;Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.; A;Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
A;Reference number: A64520; MUID:97394467; PMID:9252185  
A;Accession: G64577  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-487 <TOM>  
A;Cross-references: GB:AE000561; GB:AE000511; NID:G2313564; PIDN:AAD07525.1; PID:G2313565  
C;Superfamily: type I site-specific deoxyribonuclease chain hsdM

Query Match 6.7%; Score 165; DB 2; Length 487;  
Best Local Similarity 24.4%; Pred. No. 8.9e-05;  
Matches 63; Conservative 49; Mismatches 90; Indels 56; Gaps 12;

Qy 28 QFETPKMAATLWASMLRVDLDLGRVRLDPGAGVGSLLTAALVRLHLTERPDVAHVAV 87  
Db 208 EYTFELTASTIAIKLL-INPEFTNVKIYDPSAGTGTLALAHQIGTD----SCTLYAD 262

QY 398 NYLHESQRPLEKNVARGIMLWLNSTVLDDQ 426  
Db 469 QXGNGIQKPEPNDLNKGWMLDLG--LLDK 495

RESULT 10  
XPFS7A  
site-specific DNA-methyltransferase (adenine-specific) (EC 2.1.1.72) Paer7I - Pseudomonas  
N/Alternate names: modification methylase Paer7I  
C/Species: Pseudomonas aeruginosa  
C/Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 05-May-2000  
C/Accession: S07366  
R/Therault, G.; Roy, P.H.; Howard, K.A.; Benner, J.S.; Brooks, J.E.; Waters, A.F.; Ging  
Nucleic Acids Res. 13, 8441-8461, 1985  
A/Title: Nucleotide sequence of the Paer7 restriction/modification system and partial ch  
A/Reference number: S07366; MUID:86093653; PMID:3001639  
A/Accession: S07366  
A/Molecule type: DNA  
A/Residues: 1-531 <THE>  
A/Cross-references: EMBL:X03274; NID:G45385; PIDN:CAA27025.1; PID:G581435  
C/Genetics:  
A/Gene: paer7IM  
A/Genome: plasmid  
A/Start codon: GTG  
C/Superfamily: site-specific methyltransferase (adenine-specific) Paer7I  
C/Keywords: methyltransferase; S-adenosylmethionine

Query Match 6.6%; Score 162.5; DB 1; Length 531;  
Best Local Similarity 23.2%; Pred. No. 0.00016;  
Matches 92; Conservative 58; Mismatches 142; Indels 105; Gaps 20;

QY 53 RVLDPGAGVGSITAAVLDRL-----HTRPDV--AVHVAVETDPFVWPLYRATL 100  
Db 21 RLLEPSFGGDFLLPIIQLLSAWRAARPNGTEVDDLGLDAIRAVELHHDTPFRSTV--AAV 78  
QY 101 EECRNAYGISYD-----LVEGYLLNQGAKLDGPDFLIVIANPPY-----GKLAS 144  
Db 79 VALLKREGLSANAATALADRWLSQGDFLL--APLEGQDFVGVNPPYVRPELIPAPLLA 135  
QY 145 DSLARLATTARAVDVENVVAVFWRAVISLKEQGRGVFIVPRSWANGPYVRQFRHLMTA 204  
Db 136 EYRSRYQT--MYDRADIYPIFIERSITALSAGNLGFCADRWKMKRYGGPLR-----S 187  
QY 205 VSLDLHVFESTKVF---ADTKVKQENIVAFSVRPQSSVVLRS-----VAHGE-- 254  
Db 188 LVAREFHL-----KYVDMVDTPAFSDVIAYPAI-----TIISREGGGATRIAHRPST 236  
QY 255 -----SIASSVPFSAIVHD-----EDDDKIVHFAESASVPSAARFTL----- 291  
Db 237 DRATLITLAGLSAFTLPKADGFPVRELARVTNGAERGCSLLTRWRLFAVWRARSHCSKR 296  
QY 292 --ADLIGIGVSTGKVVDFRNRQYLTD----NLDASGVVPMVYQSNIRSGKIDWPQVGARKP 345  
Db 297 LGARFGIGVATG-----ADKAEFIGDPESLDVEPDRKLPLVTTKIDIMTGEVQWRQGVINP 351  
QY 346 QGFVAVEDVALRQLLPQGSYVVVVKQTAKEDRRRVTA 382  
Db 352 ----FAESGGLVDL---GEYPLRLRYL--EARRDVIA 379

RESULT 11  
H71935  
type I restriction enzyme (modification chain) - Helicobacter pylori (strain J99)  
C/Species: Helicobacter pylori  
A/Variety: strain J99  
C/Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 17-Mar-2000  
C/Accession: H71935  
R/Alm, R.A.; Ling, R.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;  
Ives, C.; Gibson, R.; Merberg, D.; Malls, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
Nature 397, 176-180, 1999  
A/Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
A/Reference number: A71800; MUID:99120557; PMID:9923682

```
QY 148 --ARLATTARAVDPNVVAVFWRVAVISLKEQGRGVFVPRSWANGPYRFRHWMAT-- 203
Db 159 KYRRYTT--IYRADLVVPFIQHSLELLSQGILGICSDRFYKNRYGKKLRKEITDNY 216
QY 204 ---AVSLDILHVPESRTKVF-ADTKVQENNVAVFSVRPOSSSVLRSVAHGEESI-- 256
Db 217 KVRIVDLKTKTSPFENEVTAIPAIVIKTKNVKSV-VRAVTEVITSKVQDADKDFLLS 275
QY 257 ---ASSVPSALVHDE---DDKIVHFPAESAVPSAA--RFTLAD-----LGIGV 298
Db 276 NQKPDQSSKEMKTYVFSWFAGDEPWIIQSQECREILRLRLENRFLIEDDVHSCCKIRIGV 335
QY 299 STGKVVDFERNQYLTD---NLDASGVVPMVYQSNIRSGKIDWPQVGARKP-----QGFWA 350
Db 336 ATG-----ADKVIYVDPOVDLEPVLPLVTADISSGRIIWSKXVINPENSOGGLIN 390
QY 351 VEDV-ALRQLLPQGSVVVVKQTAKED-----RRRVIAAVWDCASRVA 392
Db 391 LDDFPLKTYFQOHEEIIKKNRVAKKNPSQWPRTTIDRIYPEIVHQPKLLIPDMKNTNHIV 450
QY 393 LDNTKNYHESORPL-----EKNVARGMLMLNSTVLDQYFRAFSGHTQVNAAGDLR---- 443
Db 451 KDDGAFYPHNLYIYLPGNWNIDILRAILL---SSVK--FFIWSYATKMRGDTLRYOAO 505
QY 444 -----RLPFLCREDLILLAKVVDGLPD-----OETLDVAVARLF 478
Db 506 YLRKIRLP-----DPKSLTNDQKERLMDERVIOQSEYLDLSIVAEIY 546

RESULT 13
I40371
methyltransferase - Bacillus stearothermophilus (fragment)
C:Species: Bacillus stearothermophilus
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 26-May-2000
C:Accession: I40371
R:Rina, M.; Markaki, M.; Bouriotis, V.
Gene 150, 71-73, 1994
A:Title: Sequence of the cloned bseCIM gene: M.BseCI reveals high homology to M.BanIII.
A:Reference number: I40371; MUID:95047552; PMID:7959066
A:Accession: I40371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-579 <RES>
A:Cross-references: EMBL:X79509; NID:G619638; PIDN:CAA56041.1; PID:G619639
A:Gene: bseCIM
C:Superfamily: site-specific methyltransferase (adenine-specific) Paer71

Query Match 6.0%; Score 148; DB 2; Length 579;
Best Local Similarity 20.2%; Pred. No. 0.0023;
Matches 92; Conservative 74; Mismatches 162; Indels 128; Gaps 21;

QY 5 ASTETRQOALGKLDPTQAVLGQFFTPMKAATLMASMLRVDDLRG-----TVRVLDPGAG 60
Db 7 ANTVSQKAT-----GAHFTPDKLAIEVAKRI-LDYFKGKKNRVINVLDPACG 53

QY 61 VGSLLTAALVDRLHTEPRDVAVHVAVETDPFVVPYLRATLEECRNAYGSIYDLVEGDYL- 119
Db 54 DGEILLAI--NKVAQSMNTQLELIGVDFD---IDAINIANERLSRSGHKNFRLINKDELE 108

QY 120 -LNOQAKLD-----GPFDLVIANPPYK--LASDSLARLATTARAVDVPNVVYAFVW 168
Db 109 MYSEGNNDYDLNTELEPVDVILIINAPPYVFTQILGAEKAKLKEKNLKGKRVLDYQAFV 168

QY 169 RAVISLKEQG-RGVFIVPR--SWANGPYRFRHWMATVSLDILHVPESRTKVPADTKV 225
Db 169 AMTQQLKSGNIIIGVITSNRYLTTKGGGSTRK-----LVSNFNILIMDLGDSKFFFAAV 223

QY 226 -----KQENVIVAFSVRPOSSSVLRSVAHGEESTASSVPSALVHDE 269
Db 224 LPAIFFGKKKQKQKNSNVKPFKIYEQSDI-----EASSVNSFNSLI--- 270

QY 270 DDKIVHFAESAVPSAARFTLADLGIGVSTGKVDVFRNQ-----YLTNLI 316
```

```
Db 271 -----ELLEVNKSLGYVEDKTKYSISLGIKSIIPENYKEPWLATFEDSYEFMKNVQ 321
QY 317 DASGVPMVYQSNIRSG-----KIDWPQVGARKPQGVAVEDVALRQLLPQGSYV 366
Db 322 NAYGFIEDF--AHVKVGIKTTADSVFRISDNGEL-----PEEQIP-EDKLLRPIISADQ-- 372
QY 367 VVKEQTAKEDRRRVIAAVWDCASRVALDNKNTNYLHE 402
Db 373 -----ANKW-SVSLVGNKKVLYTHE 392
```

## RESULT 14

F64633

site-specific DNA-methyltransferase (EC 2.1.1.1.-) HP0910 - Helicobacter pylori (strain 26)

C:Species: Helicobacter pylori

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Aug-2000

C:Accession: F64633

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.; Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: F64633

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-379 <TOM>

A:Cross-references: GB:AE000600; GB:AE000511; NID:G2314042; PIDN:AAD07955.1; PID:G231404

C:Genetics:

A:Start codon: TTG

C:Superfamily: site-specific DNA-methyltransferase HP0910

C:Keywords: methyltransferase; restriction modification system

Query Match 6.0%; Score 147; DB 1; Length 379;

Best Local Similarity 28.0%; Pred. No. 0.0015;

Matches 60; Conservative 28; Mismatches 76; Indels 50; Gaps 11;

QY 15 LGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRGTVRVLDPGAGVGSILTAALVDRLHT 74

Db 1 MNLDIKT---LGQVFTPKK---IVDFMLTLKHNGS--VLFPSSAGDSGSLKEL----- 46

QY 75 ERPDVAVHVAVETDPFVVPYLRATLEECRNAYGSIYDLVEGDYLLNQAKLDGPPDLVI 134

Db 47 -----KKAVRIEIDPKICP-----KNALCMDF----FDY-----PLENQPDITII 81

QY 135 ANPPYK---LASDSLARLATTARAVDVPNVVAVFWRVAVISLKEQGRGVFIVPRSWANG 191

Db 82 GNPPYVKHDIAPSTKEKLHYSL-FDERSNLYFFIEKAIKHLKPKGELIFITPDLFLKS 140

QY 192 PYRQPRHWMATVSLDILHVPESRTKVPADTKV 225

Db 141 TSSVKLNEWIYKEGT--ITHFFE-----LGQDKV 167

## RESULT 15

H97333

site-specific modification DNA-methyltransferase [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C>Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 01-Mar-2002

C:Accession: H97333

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: H97333

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-571 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81459.1; PID:G15026627; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

Search completed: October 1, 2004, 16:11:00  
Job time : 19.8208 secs

**This Page Blank (uspto)**



```

QY 71 RLHTEPDVAVHVVAVETDPFVYVLRATLEECR---NAYGISYD-LVGEVDYLLNQAK 125
Db 78 NA-TLNGAKSISSTCYEISEVMVYNLIQVLDLCKIRAMEFEVNMWOCKIIESDFQASVEQ 136
QY 126 L-----DGPFDLVANPPYKGLASDLARLATTARAVDPVNPVYAFWVRVVISLKEQGRG 180
Db 137 LIENYSKYKALINPPYKIAKGRREALLOKVGIEASNLISAFVALAIKQLKGGEL 196
QY 181 VFIIVRSWANGPYRQFRHMLMTAVSLDILHVFESRTKVFACTKVQENVIIVAFSVRQFS 240
Db 197 VAITPRSCFNGFYFNDKQMLDECSLNKIHVNSRKSFAKADNVLQENIYIHLTKGETQ 256
QY 241 SSV--VLSRVAHGRESTASSVPFSALVHDEDDDKIVHFA-----ESASVPSAARFTLA 292
Db 257 RKWTVYSTCANDINPTIFEVFPDEIVKSNPNPDLFIHVTNEQRELANKAGGLPCSL 316
QY 293 DLGIGVSTGKVVDPRNRVLTNDLDSGVVPMVYQSNIRSGKIDWPQVGARKPQGFVAVE 352
Db 317 DLGIGVSTGKVVDPRNRVLTNDLDSGVVPMVYQSNIRSGKIDWPQVGARKPQGFVAVE 374
QY 353 DVALRQLLPQGSYVVVVKQTAKEDRRRRIAIVW----DGASVALDNKTNVHESORPLE 408
Db 375 EATNLMVNGIYVLTRELTAKERRIVASIIYYPDIANVDTVGFDNKINYPHANGKPLD 434
QY 409 KNAVGLMLNLTVDLDOYFRAFSGHTQVAGDLRLPFLCREDLILLAKVVPDGLPQOE 468
Db 435 ISLAKGLMWFLNSTLIDKYFROMNGHTQVATDLRALRPTREQLIEDIANQVDFEFGT 494
QY 469 TLDVAV 474
Db 495 KIDEII 500

RESULT 2
MTTB BACSU
ID MTTB BACSU STANDARD; PRT; 501 AA.
AC P33563;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Modification methylase BsuBI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BsuBI) (M.BsuBI).
GN HSDBM OR HSBM
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISB8;
RX MEDLINE=93126092; PubMed=1480472;
RA Xu G.-L., Kapfer W., Walter J., Trautner T.A.;
RT "BsuBI -- an isospecific restriction and modification system of PstI:
RT characterization of the BsuBI genes and enzymes.";
RL Nucleic Acids Res. 20:6517-6523(1992).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CTGACG, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRAND, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BsuBI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L01541; AAA18169.1; -.
CC DR PIR; S35515; S35515.
CC DR REBASE; 3336; M.BsuBI.
CC DR InterPro; IPR002296; N12N6 mtfase.
CC DR InterPro; IPR002052; N6_Mtase.

```

```

DR PRINTS; PR00507; N12N6MTFRASE.
KW PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 501 AA; 57197 MW; AC33B2944699653E CRC64;

Query Match 26.8%; Score 662.5; DB 1; Length 501;
Best Local Similarity 36.4%; Pred. No. 6.1e-41;
Matches 173; Conservative 80; Mismatches 193; Indels 29; Gaps 12;

QY 26 LQGFPTPKAATLMASMLRVDDLRCTVAVLPAGVGAGVSGISLTAALVDRLHTEPDVA-VHVV 84
Db 25 LQGFPTPSISIFWACLFSDEKL--NNAKVLDAAGAGISLTAALVISENIGKADLHL- 82
QY 85 AVETDEFFVYVLRATLEECRNAYGISYDVLVEGDY-----LLNQG----AKLDGPFDLVI 134
Db 83 -LEIDEMLEPYLSETLALFKDYIEINSQIIDDFTIEMAAYSLLDEESLLAKQKQRFTHAI 141
QY 135 ANPPYKGLASDLARLATTARAVDPVNPVYAFWVRVVISLKEQGRGVFIIVRSWANGPY 194
Db 142 LNPPYKIKSNSKHKRLKAKGIEITNLYSAFVALTVDLMSDGGIEIVFIIPSCFNGPYV 201
QY 195 RQFRHMLMTAVSLDILHVFESRTKVFACTKVQENVIIVAFSVRQSSSVLSRSVAHGSE 254
Db 202 RFRQHLNKTSLIKHMLFESRDKAFKQDEVLUQENVISKLEKGTQVEDVKISITDDSF 261
QY 255 SIAS--SVFPFSALVHDEDDDKIVHFAESASVPSAAR-----FTLADLIGVSTGKVVDPRN 308
Db 262 VTRSYRYFEKIVQPNDEIKFTHINTNEETLIEKHPNVCSLEELNIEVSTGVPVDPFRV 321
QY 309 ROYLTNDLDSGVVPMVYQSNIRSGKIDWPQVGARKPQGFVAEDVALRQLLPQGSYVVV 368
Db 322 KENLRE-MPEGCTVPLFYPNHFVGTSLSEYPKM-MKKPNAILIRNEKVE-KWLYPENGHYVVV 378
QY 369 KQQTAKEDRRRIAIVWDCAS----RVALDNKTNVHESORPLEKNVAGSLMLNLTVDL 424
Db 379 KFPSKKEKRIRVAGVLTPEVNDPVVGFENGLNVLHYNKSGISKEVAGLYAYLNSTPV 438
QY 425 DOYFRAFSGHTQVAGDLRLPFLCREDLILLAKVVPDGLPD--QETLDVAVVARL 477
Db 439 DKYFRIFNGHTQVATDLRTMKFSPRDLISLGKWIENIVNGVQVEIDSKLEEL 493

RESULT 3
MTTB THEAQ
ID MTTA THEAQ STANDARD; PRT; 421 AA.
AC P14385;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Modification methylase TagI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase TagI) (M.TagI).
GN TAQIM.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OX Thermus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YTI1;
RX MEDLINE=88096538; PubMed=2827113;
RA Slatko B.E., Benner J.S., Moran L.S., Jager-Quinton T., Simcox T.G.,
RA van Cott E.M., Wilson G.G.;
RT "Cloning, sequencing and expression of the Tag I restriction-
RT modification system.";
RL Nucleic Acids Res. 15:9781-9796(1987).
RN [2]
RP REVISIONS.
RC STRAIN=YTI1;
RX MEDLINE=92201707; PubMed=1551602;
RA Barany F., Slatko B., Danzitz M., Cowburn D., Schildkraut I.,
RA Wilson G.G.;
RT "The corrected nucleotide sequences of the TagI restriction and
RT modification enzymes reveal a thirteen-codon overlap.";

```

RL Gene 112:91-95(1992).

RN [3]

RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).

RX MEDLINE=95062184; PubMed=7971991;

RA Labahn J., Granzin J., Schluckebier G., Robinson D.P., Jack W.E.,

RA Schildkraut I., Saenger W.;

RT "three-dimensional structure of the adenine-specific DNA

RT methyltransferase M.Taq I in complex with the cofactor S-

RT adenosylmethionine.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:10957-10961(1994).

CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE

CC TCGA, CAUSES SPECIFIC METHYLATION ON A-4 ON BOTH STRANDS AND

CC PROTECTS THE DNA FROM CLEAVAGE BY THE TAAI ENDONUCLEASE.

CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-

CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M76681; AAA27506.1; -

DR EMBL; Y00499; CAA68551.1; ALT\_SEQ.

DR PIR; JN0257; JN0257.

DR PDB; 2ADW; 27-JAN-97.

DR PDB; 1AQL; 12-FEB-97.

DR PDB; 1AQL; 12-FEB-97.

DR PDB; 1G38; 12-MAR-01.

DR REBASE; 3520; M.TaqI.

DR InterPro; IPR002296; N12N6 mtfrase.

DR InterPro; IPR002052; N6 Mtase.

DR InterPro; IPR000051; SAM bind.

DR PRINTS; PR00507; N12N6MTFRASE.

DR PROSITE; PS00092; N6\_MTASE; 1.

KW Transferase; Methyltransferase; Restriction system; 3D-structure.

FT HELIX 25 34

FT TURN 39 40

FT STRAND 42 46

FT TURN 47 47

FT TURN 49 50

FT HELIX 52 61

FT STRAND 66 71

FT TURN 74 76

FT TURN 81 82

FT STRAND 84 87

FT TURN 90 92

FT STRAND 99 104

FT HELIX 122 132

FT TURN 134 135

FT TURN 138 139

FT HELIX 142 152

FT TURN 153 153

FT STRAND 154 165

FT HELIX 166 169

FT TURN 170 170

FT HELIX 172 174

FT HELIX 175 184

FT STRAND 186 193

FT TURN 199 200

FT STRAND 203 209

FT STRAND 215 223

FT TURN 224 225

FT STRAND 226 235

FT TURN 236 237

FT HELIX 248 255

FT TURN 256 256

FT STRAND 257 259

FT HELIX 260 263

FT STRAND 264 264

FT HELIX 272 276

FT TURN 277 277

FT TURN 279 280

FT STRAND 282 282

FT TURN 287 288

FT STRAND 289 291

FT STRAND 293 293

FT HELIX 295 297

FT STRAND 298 299

FT TURN 300 301

FT STRAND 302 303

FT STRAND 313 315

FT HELIX 316 321

FT TURN 322 322

FT HELIX 324 327

FT STRAND 330 333

FT TURN 343 346

FT STRAND 347 348

FT STRAND 352 352

FT STRAND 356 360

FT TURN 362 363

FT STRAND 364 366

FT HELIX 368 375

FT TURN 376 376

FT HELIX 378 388

FT TURN 389 390

FT HELIX 397 400

FT TURN 401 402

FT STRAND 404 406

FT TURN 407 409

FT STRAND 410 412

SQ SEQUENCE 421 AA; 47848 MW; 6F7D2081EB7FCE45 CRC64;

Query Match 7.4%; Score 183; DB 1; Length 421;

Best Local Similarity 23.7%; Pred. No. 5.1e-06;

Matches 108; Conservative 58; Mismatches 187; Indels 102; Gaps 20;

QY 26 LGQFTPMKAATMASMLRVDDLGRGTVRLDPCGAGVCSLAALVDRLHTEPDPVAVHVA 85

DB 18 LGRVETPEVVDVFWSLAEAP--RGG-RVLEPCAHGPF--RAFREAHGTGYRFVG 69

QY 86 VETDPFV--PYIRATLEECRNAYGISYDLVEGYLLNQAKIDGPFDLVIANPPYG-- 140

DB 70 VEIDPKALDPPWAEGIL-----ADFLWEPGE---AFDLILGNPPYGIV 111

QY 141 -----KLASLARLATARADVDPNVYVAFWRAVISLKEQGRGVFIVPRSWA 189

DB 112 GEASKYPIHVFKAVKDYKKAFTWKG--KYNLYGAPLEKAVRLKPGGVLFVVPATWL 169

QY 190 NGPYRQFRHMLMTAVSLDILHVFESRTKYFADTKYQENVIVAFSVRPOSSSVLSRSV 249

DB 170 VLEDFAALLREFLAR-----EGKTSVYVYIGEVFPQKKVSAVIRFOKSGKLSLD 219

QY 250 AHGEESIASVPFSAVHDEDDDKIVHF-AESASVPSAARFTLADLIGVSTGKVVDFRN 308

DB 220 TQESSEGFTLLWAEYPHWEGE--IIRFETEETKLEISGMPGLDLFHPFAARSPEFKK 277

QY 309 RQVLTNLDASGVVPMVYOSNIRSGKIDW-----POVGARKPGQFVAVEDVALRQL 359

DB 278 HPAVRKE-PQGLVPLVTGRNLKPGWVDYKKNHSGLWMPKERAKELRDFVAT----- 328

QY 360 LPOGSSVYVVKROTAKEDRRRIVAAVWDGASRVALDNKTNVYLHESQRPLEKNVAR----GL 415

DB 329 ----PHLVV-----AHTKGRVVAA--W-----DERAYFWREBFHLLKPEGRVLDPSLL 371

QY 416 MLWLNSTVLDQY----FRAFSGHTQVNAAGDLRLRP 446

DB 372 VQWLNSEANQKHVRTLYRDFVPHLTIRM--LERLP 404

RESULT 4

MTT8 THETH

ID MTT8 THETH

AC P29749; STANDARD; PRT; 428 AA.

DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 30-MAY-2000 (Rel. 33, Last annotation update)  
DE Modification methylase TthHB81 (EC 2.1.1.72) (Adenine-specific  
DE methyltransferase TthHB81) (M.TthHB81).  
GN TTHB81M.  
OS Thermus thermophilus.  
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;  
OC Thermus.  
OX NCBI\_TaxID=274;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HB8 / ATCC 27634;  
RX MEDLINE=92201699; PubMed=1339363;  
RA Barany F., Danzitz M., Zebala J., Mayer A.;  
RT "Cloning and sequencing of genes encoding the TthHB81 restriction and  
RT modification enzymes: comparison with the isoschizomeric TaqI  
RT enzymes."  
RL Gene 112:3-12(1992).  
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
CC TCGA, CAUSES SPECIFIC METHYLATION ON A-4 ON BOTH STRANDS AND  
CC PROTECTS THE DNA FROM CLEAVAGE BY THE TTHB81 ENDONUCLEASE.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M74795; AAA27488.1; -.  
DR PIR; JH0634; JH0634.  
DR HSSP; P14385; 2ADM.  
DR REBASE; 3523; M.TthHB81.  
DR InterPro; IPR002296; N12N6\_mtfase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR InterPro; IPR000051; SAM\_bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
KW Transferase; Methyltransferase; Restriction system.  
SQ SEQUENCE 428 AA; 48317 MW; 4CDFB8B1959EF5F6 CRC64;  
  
Query Match 7.0%; Score 172.5; DB 1; Length 428;  
Best Local Similarity 22.3%; Pred. No. 3.1e-05;  
Matches 109; Conservative 57; Mismatches 174; Indels 149; Gaps 21;  
  
QY 18 LDPTTQA-VLGGFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVSLTAALVDRLHTER 76  
DB 7 LPTTAGRSIGRVETP---PGLVRFVWGVAEARKGVRLVLEPCADGFFL-----RAPPEA 58  
  
QY 77 PDVAVHVAVETDPFVVV---PYLRATLEECRNAYGISYDLVEGDYLLNQAKLDGDFDLV 133  
DB 59 HGTYRFGVGVIEDPHALDLPMAEG-----VVADFLLWEPGE---AFDLI 100  
  
QY 134 IANPPYKGLASDSLARI-----ATTARADVNVVYAFVFWRAVISLKEQGRGVF 182  
DB 101 LGNPPYGIYGEASKYPIHVLREYKGLYKTLSTWKGKYNLYGAFIEKSVRLLEGGTLVF 160  
  
QY 183 IYPRSVANGPYRQFRHMLMTAVSLDILHVFESRTKVFADTKVQENIVVAFSVRPQSSS 242  
DB 161 VPDATVILVLDLDFSLRSFLAREGTEVYILGE---VFPGRKYSA--VVLRF--RKGKG 212  
  
QY 243 VVLSRSVAHGE-----ESTASSVPFSALVHDEDDDKIVH 276  
DB 213 LALWDFTRDGETFTPLLWSEKPGMEGRDHRFETGWTREMEASGPPLGSLFH----- 263  
  
QY 277 FAESASVPSAARTLADLGLGVSTGVKVVDFERNQYLTDLNLDASGVPMVQSNIRSKID 336  
DB 264 -----IPFAARGP-----EPKKGHFAVQKEPE--PGLVPVLTGRNLKPGWID 302

QY 337 W-----POVGARKEQGFVAVEDVALRQLLPGSGVYVVKRQTAKEDRRRVIAAVWDG 387  
DB 303 YESNHSLWMPKERAKELRDFYATPHL-----VWARTKGTK-----VVAAM-- 343  
  
QY 388 ASRVALDNKTYLHESQRPLEKNVAR-----GLMLWLNSTVLDOY-----FRAPSGHTQVNA 439  
DB 344 -----DEKAYPWRREEFHLPLKPGVELDPLFLVEWLNSDKIQEYVKTLYRDPVPHLRLM 397  
OX 440 GDLRRRLPEL 448  
DB 398 --LERIPAL 404

## RESULT 5

MTSL STRAL STANDARD; PRT; 587 AA.  
AC Q53609;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Modification methylase Sali (EC 2.1.1.72) (Adenine-specific  
DE methyltransferase Sali) (M.Sali).  
GN SALIM.  
OS Streptomyces albus G.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1962;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95129852; PubMed=7828868;  
RA Rodicio M.R., Quinton-Jager T., Moran L.S., Slatko B.E., Wilson G.G.;  
RT "Organization and sequence of the Sali restriction-modification  
RT system."  
RL Gene 151:167-172(1994).  
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE  
CC GTGAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND  
CC PROTECTS THE DNA FROM CLEAVAGE BY THE SALI ENDONUCLEASE.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-  
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U01232; AAA81887.1; -.  
DR HSSP; P14385; 2ADM.  
DR REBASE; 3491; M.Sali.  
DR InterPro; IPR002296; N12N6\_mtfase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR InterPro; IPR000051; SAM\_bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
KW Transferase; Methyltransferase; Restriction system.  
SQ SEQUENCE 587 AA; 64975 MW; B482F374397FF999 CRC64;

Query Match 6.9%; Score 171.5; DB 1; Length 587;  
Best Local Similarity 22.6%; Pred. No. 5.6e-05;  
Matches 122; Conservative 80; Mismatches 187; Indels 151; Gaps 28;  
  
QY 27 GQFPTPMKAATLMASMLRVLDLRGTVRV--LDPGAGVSLTAALVDRLHTERPDVAVHV 83  
DB 32 GQFATP---PSLAGEIMRYTIDLHEETRNFLNFPSCGSGSFFSALLRNLGDKKIE---HA 85  
  
QY 84 VAVETDPFVVPYIRATLEECRNAYGISYDLVEGDYLLNQAKLDGDF--DLVIANPPYK- 141  
DB 86 VGVELDP-----RFSKAASDLMTGQGLRVIEGDF--TSPSLVSGPVASLLVANPPVVRH 137  
  
QY 142 --LASDLSARLATTAARVD-----VFNVYVAFVFWRAVISLKEQGRGVFIIPRSWANG 191

```

Db 138 HHLGIDQKRD--VARCADQIGKPSGUSGLYLYFVLLSHLLRADAVSTWLPSEFMDV 195
Qy 192 PYRQFRHLMTAVSLDILHVFESTKVFADTKVKQENVI VAFSVRPOSSSVLSRS--- 248
Db 196 NYGTALKEYLATRVOLVRIHQVDAEVOFDDALV-----TSSVVVFRNSPP 241
Qy 249 -VAH-GEBSIASVPFSAVHDEDDDKIVHFAESASVPSAARFTL-----ADLIGIVS 299
Db 242 RPKHTABFSFGTL-----SEPKVTHQIPSAALTPKAKSRVYGVMPADINLKQT 292
Qy 300 TKGVDV-PRNRQYLTDNLDSGVVPMVYQSN-----RSKID----- 336
Db 293 GPKLSDFFKIRGLATNSAFIIPRSEARLGIKRNFLRPLPSRKLKGDATADASG 352
Qy 337 WPOVGAR-----KPOGFVAVEDVALRQLLP-----QGSYVVVKRT-AKEDRRR--V 380
Db 353 WPDPEQALLDCPLIEDLLENPALAAYLSTADEKIRGGYLVSKSPWTKQOREPAP 412
Qy 381 IAAVWDGASRVLDNKNTYHLESORPLE--KN-----VARGMLMLNSTVLQDYFRAFSGH 434
Db 413 ILTYMGRGK-----DDQHLRFIRNDSDAVATNMVLMYPTALLQRYL----- 456
Qy 435 TQVWAGDLRLPLFCREDLILLLA-----XVVDGGLPDQF-----TLDVV 474
Db 457 ----AGDPERIKQVHKALLAITAADLRGGGRVYGGGLHKMEPKELAAPADGIATLDPVL 512

```

## RESULT 6

```

MTC3 CHVN1
ID MTC3 CHVN1 STANDARD; PRT; 377 AA.
AC P10835;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Modification methylase CviBIII (EC 2.1.1.72) (Adenine-specific
DE methyltransferase CviBIII) (M.CviBIII).
GN CviBIII.
OS Chlorella virus NC-1A (CV-NC1A).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10504;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88096540; PubMed=3320956;
RA Narva K.E., Wendell D.L., Skrdla M.P., van Etten J.L.;
RT "Molecular cloning and characterization of the gene encoding the DNA
RT methyltransferase, M.CviBIII, from Chlorella virus NC-1A.";
RL Nucleic Acids Res. 15:9807-9823(1987)
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC TCGA, CAUSES SPECIFIC METHYLATION ON A-4 ON BOTH STRANDS.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06618; CAA29835.1; -.
CC HSSP; P14385; 1AQQ.
CC REBASE; 752; M.CviBIII.
CC InterPro; IPR002296; N12N6_mtfrase.
CC InterPro; IPR002052; N6_Mtase.
CC PRINTS; PR00507; N12N6MTFRASE.
CC PROSITE; PS00092; N6_MTASE; 1.
CC Transference; Methyltransferase.
CC SEQUENCE 377 AA; 42895 MW; B22D471367D61BBC CRC64;

```

Query Match

6.9%; Score 169.5; DB 1; Length 377;

Best Local Similarity 23.7%; Pred. No. 4.3e-05;

```

Matches 75; Conservative 46; Mismatches 123; Indels 73; Gaps 16;
Qy 27 GQFTPKMAATLMASMLRVDDLRTGTVRLVDPGAGVGSITAAALVDRLHTERPDVAHVAV 86
Db 26 GIFFTPKTVREKLFQFTEHFQNTPGFSILEPSCGTGEIISECVERF-----PLASIKGVEL 81
Qy 87 ETDPFVVPYLRATLEECRNAGV-LSYDLVDEGDYLLNOGAKLDGPDLDVIANPPY-----G 140
Db 82 DNDMSTI-----CSKYAEYNVDIVNEDFLLWK-----GRFDFIVGVNPPYVVRPSG 128
Qy 141 KLASDSLARLATTARAVDVNPNVYVAVVRAVIS-LKEQGRGVFIVPRSWANGPYVVRQFRH 199
Db 129 YKDNRIAK-----GRS-----NLYVEFLYKCI TEHLKEDGILAFIIPSTIGNSSEYEPK 180
Qy 200 WLMTAVSLDILHVFESTK-VFADTKVKQENVI VAFSVRPOSSSVLSRSVAHGEESTAS 258
Db 181 LIIT--LDILS-FEILDKHD FCDTNR-----LCSIVIKNSPGTKYT--- 220
Qy 259 SVFPSALVHDEDDDKIVHFAESASVPSAARFTLADLGGVSTGKVVDFRNRQYLTDLNDA 318
Db 221 ---YRDYICXD-----IPHGNSYIGSLDKFKTGFAMANVKN-FTFDKSE- 263
Qy 319 SGVVPVYQSNIRSGKI 335
Db 264 ---IPFTSSNIKINEI 277

```

## RESULT 7

```

MTAL ACICA
ID MTAL ACICA STANDARD; PRT; 540 AA.
AC P25201;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase AccI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase AccI) (M.acci).
GN ACCIM.
OS Acinetobacter calcoaceticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Moraxellaceae; Acinetobacter.
OX NCBI_TaxID=471;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 49823;
RX MEDLINE=91345839; PubMed=1368703;
RA Kawakami B., Hilzheber C., Nagatomo M., Oka M.;
RT "Cloning and nucleotide sequences of the AccI restriction-modification
RT genes in Acinetobacter calcoaceticus.";
RL Agric. Biol. Chem. 55:1553-1559(1991).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GTMKAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE ACCI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -!- SUBUNIT: Monomer.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10671; BAA01523.1; -.
CC REBASE; 3271; M.Acci.
CC InterPro; IPR002296; N12N6_mtfrase.
CC InterPro; IPR002052; N6_Mtase.
CC PRINTS; PR00507; N12N6MTFRASE.
CC PROSITE; PS00092; N6_MTASE; 1.
CC Transference; Methyltransferase; Restriction system.
CC SEQUENCE 540 AA; 63092 MW; A2569693712E0F59 CRC64;

```



```

RC STRAIN=V;
RX MEDLINE=93380657; PubMed=8370531;
RA Gonzalez E., Vasquez C.;
RT "Characterization of the bstVIRM genes encoding the Bacillus
RL stearothermophilus v restriction-modification system.";
Gene 131:103-106(1993).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC CTGACG, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BSTVI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L07642; AAA51408.1; ..
DR PIR; JN0797; JN0797.
DR HSSP; PL4385; 2ADM.
DR REBASE; 3331; M.BstVI.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Methyltransferase; Restriction system.
SQ SEQUENCE 561 AA; 65702 MW; 4D463062E4941859 CRC64;

Query Match          6.1%; Score 151; DB 1; Length 561;
Best Local Similarity 21.3%; Pred. No. 0.0016;
Matches 112; Conservative 85; Mismatches 211; Indels 118; Gaps 25;

QY   53 RVLDPGAGVGLTAALVDRLHTE-----RPDVAVH-----VVAVETDFVPWPLYRATLE 101
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   39 KLDPSPFGDGVFLEAAVHRLMDSLIIRGYRPNELIDHLGNCIRGLELREAYAQGRHLQ 98
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   102 EC RNAYGISVD-----LVGEDYLLNQAKLDG--PDFLVTANPPY--GKLASDL-- 147
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   99 KVLGYGFSKPEINWLNIQWIQADEFVVGNEPYVRQELIQDELIK 158
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   148 --ARLATTAARAADVNPVYAFWRVAVLSLKQEGGVFIVPSRWANGPYRQRFHWLMT-- 203
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   159 KYRKRYTT--IYDRADLYVEFFIOHSLELSEQTLGIICSDRFTKNRYGKKURKFITDY 216
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   204 ----AVSLDILHFVESKTGF--ADTKVKQENVIAFSVRPQSSSVLSRSVAHEESI-- 256
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   217 KVRYIVDLHKTSFPENEVTPAIYVLTKNYDKSV-VRAVYTEVITSKVCDQAKDFLLS 275
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   257 ----ASSVPFSALVDE----DDDKIVFHAEASVPSAA--RFTLAD-----LGIGV 298
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   276 NKPDQSSKBMKTYVSEWFAGDEPIIIQSCEIRELRLENRFLIEDDVHSCIRIGV 335
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   299 STGWVDFNRQYLTD---NLDSAGVWPVMVYQSNIRSGKIDWPQVGARKP-----QGFA 350
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   336 ATG ----ADKVIVDPQQVDIEPFVELLEVTTADISSGRIIWSGKHVINPFNSDGLIN 390
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   351 VEDV-ALRQLLPQGSYYVVKRQTAKED-----RRRVAAVWDGASRVA 392
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   391 LDFFPLKTVFQHEELIKRNNAVKNPQSFQFRTIDRIYPEIVHQPKLLIPDMKNTNHIV 450
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   393 LDNKTNVYHESQRPL-----EKNVARGLMLNWSTVLDQYFRAFSGHTOVNAGDLR---- 443
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   451 KDGAFYPHNLYXYILPGNNWIDILRAILL---SSVVK--FTWSYATKMGDTLRYQAQ 505
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

QY   444 ----RLPFLCREDLLILAKVDPDGLPD-----QETLDVAVLARLF 478
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
QY   506 YLRKIRLP-----DPKSLTNDQKERLMDSERVISQOEYLDYSIAEIY 546
Db   :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

RESULT 10

```

RESULT 10

```

MTC1_BACST
ID MTC1_BACST STANDARD; PRT; 579 AA.
AC P43423;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase BseCI (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BseCI) (M.BseCI).
GN BSECI.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=142;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95047552; PubMed=7959066;
RA Rina M., Markaki M., Bouricots V.;
RT "Sequence of the cloned bseCIM gene: M.BseCI reveals high homology to
RT M.BanIII.";
RL Gene 150.71-73(1994).
CC
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC ATCGAT, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BANIII ENDONUCLEASE.
CC
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ---
CC EMBL; X79509; CAA56041.1; -.
CC PIR; I40371; I40371.
CC DR HSSP; P14385; IAQJ.
CC
CC REBASE; 3536; M.BseCI.
CC DR InterPro; IPR002296; N12N6_mtfase.
CC DR InterPro; IPR002052; N6_Mcase.
CC DR InterPro; IPR000051; SAM_bind.
CC DR PRINTS; PR00507; N12NGMTFRASE.
CC DR PROSITE; PS00092; N6_MTASE; 1.
CC KW Transferase; Methyltransferase; Restriction system.
CC SQ SEQUENCE 579 AA; 66774 MW; A4805FBEBA01835C CRC64;

Query Match 6.0%; Score 148; DB 1; Length 579;
Best Local Similarity 20.2%; Pred. No. 0.0028;
Matches 92; Conservative 74; Mismatches 162; Indels 128; Gaps 21;

QY 5 ASTETTRQALGKGLDPTTQAVLGQFTPKAAATLMASMLRVDLDLG-----TVRVLDPCAG 60
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 61 VGSLLTAALVRLHTEPRDVAHVHVAVETDFVVPYLRATLEECRNAYGISYDLVEGDL- 119
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 54 DGBELLMLAI--NKVAQSMNIQLELIGVDFD--IDAINIANERLSRSHKFNRLINKDFLE 108
QY 120 -LNQGAKL-----GPDFLVIANDPYGK--LASDSIARLATATTARVDVPNVYVAFV 168
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 109 MVSSEGNDFLNFIELEPVDIIITIANPPYVRTQILGAEKQAQKLEKFNLRVLDYQAFLV 168
QY 169 RAVISLSKEQG-RGVFTVPR--SWANGPPYRQRFHMLTAVSLDLILHVFESRTKVADTKV 225
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 169 AMTQQLKSGNIIGVITSNRYLITKGGSTRKF-----LVSNFNILIMDLGDSKFFPEAAV 223
QY 226 -----KQENVIVAFSRPQSSSVLRSVAHGEESIASVSPFSALVHDE 269
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 224 LPAIFFGEKKNKKEYQKNSNPVKFKFIYEQSDI-----EASSSVNSFNSLI--- 270
QY 270 DDDKIVHFAESASVSPAAPTADIGIVSTGKVDVFNKQ-----YLTNML 316
QY 271 -----ELLEVKNKSGYVEDKTYISLQKIIISPENYKFWPILATEDYEFWMKVNQ 321

```

```

QY 317 DASGVVPMVYQSNIRSG-----KIDMPQVGARKPQGVAVEDVALRQLLPQGSYV 366
Db 322 NAYGFIEDF--AHVKGIGKTTADSVFIRSDWGL-----PEQIP-EDKLLRPIISADQ-- 372
QY 367 VVKRQAKEDRRRVIAAVWDGASRVALDNKNTYILHE 402
Db 373 -----ANKW-SVSLVGNKKVLYTHE 392

RESULT 11
MTX1_XANCR
ID MTX1_XANCR STANDARD; PRT; 527 AA.
AC P96188;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase Xami (EC 2.1.1.72) (Adenine-specific
DE methyltransferase Xami) (M.Xami).
GN XAMIM
OS Xanthomonas campestris (pv. amaranthicola).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=54735;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 11645;
RX MEDLINE=97276896; PubMed=9130589;
RA Gomez P., Ribas-Aparicio R.M., Pelaez A.I., Gomez A., Rodicio M.R.;
RT "Isolation and nucleotide sequence of the gene encoding the Xami DNA
RT methyltransferase of Xanthomonas campestris pv. amaranthicola.";
RL Biochim. Biophys. Acta 1351:261-266 (1997).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC GTGCAC, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE XAMI ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U77781; AAD13686.1; -
DR REBASE; 3526; M.Xami.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; FALSE_NEG.
DR TRANSFERASE; Methyltransferase; Restriction system.
KW TRANSFERASE; Methyltransferase; Restriction system.
SQ SEQUENCE 527 AA; 57535 MW; B80C43CE684D8956 CRC64;

Query Match 5.6%; Score 138.5; DB 1; Length 527;
Best Local Similarity 21.5%; Pred. No. 0.012;
Matches 111; Conservative 68; Mismatches 204; Indels 133; Gaps 25;

QY 15 LGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRGT-VRVLDPGAGVGSLLTAALVDRHL 73
Db 66 LGHRAFRAGAV---YTP---AFIVRSMTWTWLAAGQSPARIVDPGAGSGRFLAAGEAF- 117
QY 74 TERPDVAHVAVVETDPFVVVYPRATLEECRNAYGISYDLVEGDYLLNQAKLD--GPPD 131
Db 118 ---PD---AQIVAVEMDPLAALMLRANLS---ARGWTDRAITVMVKDY---REVKLPPCAGIT 167
QY 132 LVIANPPYK---LADSLARLATT-----ARAVDVPNVVAFVRAVITSLKQGRGVFI 183
Db 168 AFIGNPPYVKKHDI GEDWKAWYASNFAGYIGIKASALAGLHLHFLQTRLLAKAGDVGAFT 227
QY 184 VPRSWANGPYRQFRHLMWMTAVSLDILHVPFESRTKVFADTKVKQENVIVAFSV-----RPQ 239

```

```

Db 228 TSAEMWDVNYGSALRRLLLLDELGGIALHVLPEPTVEAPPGTATTA--AIACFRVGETARPV 285
QY 240 SSSVV-----LSRSVAHGEESIASVPPFSALVHDEDDDKIVHFAESASVPSAARETL 291
Db 286 RVRFIDELTNLNGITKGTDDIPREQLQAASRSLIVR-----PSAPAMA----- 328
QY 292 ADLGIG-----YSTGKVVDFRNRQYLTDNLDASGVVPMVYQSNIRSGKIDWFOVGARKPQG 347
Db 329 GDIELGELFRVHRGQVTG-ANGIWIAGE-HAQGLPDRVKMPAVTKAK-DLIQAQAHLSA 385
QY 348 FVAVEDVALRQLLPQGSVYVVKRQAKEDRRVIA-----AVWDGASRVALDNKNTYILHE 403
Db 386 EV-----LRVIDLPTDL-----DDFTKEERRRISFLSWAKLHGADQ-----SYIAQH 429
QY 404 QR-----PLEKNVARGLMLW 418
Db 430 RRAWWSVGLKAPAPILCTYMARPPQFTLNACDARHINIAHGLYPREPLAAGIMASLVTV 489
QY 419 LNSTVLDOYFRAFSGH-TQVNAGDLRRLPFLCREDL 453
Db 490 LNKNIINTSGRTYAGGLTKFEPKEIERLRIPSLENI 525

RESULT 12
MTB3_BACAR
ID MTB3_BACAR STANDARD; PRT; 580 AA.
AC P22772; O44656;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Modification methylase BanIII (EC 2.1.1.72) (Adenine-specific
DE methyltransferase BanIII) (M.BanIII).
GN BANIIIM.
OS Bacillus aneurinolyticus.
OC Bacteria; Firmicutes; Bacillales; Paenibacillaceae;
OC Aneurinibacillus group; Aneurinibacillus.
OX NCBI_TaxID=1391;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91242093; PubMed=1368640;
RA Kawakami B., Sasaki A., Oka M., Maekawa Y.;
RT "Nucleotide sequence of the gene coding for the BanIII DNA
RT methyltransferase in Bacillus aneurinolyticus.";
RL Agric. Biol. Chem. 54:3227-3233 (1990).
CC -!- FUNCTION: THIS METHYLASE RECOGNIZES THE DOUBLE-STRANDED SEQUENCE
CC ATCGAT, CAUSES SPECIFIC METHYLATION ON A-5 ON BOTH STRANDS, AND
CC PROTECTS THE DNA FROM CLEAVAGE BY THE BANIII ENDONUCLEASE.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA adenine = S-
CC adenosyl-L-homocysteine + DNA 6-methylaminopurine.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83417; CAA58443.1; ALT_SEQ.
DR FIR; JH0224; JH0224.
DR REBASE; 3296; M.BanIII.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
DR TRANSFERASE; Methyltransferase; Restriction system.
KW TRANSFERASE; Methyltransferase; Restriction system.
SQ SEQUENCE 580 AA; 66344 MW; 0E4DF7357EC394C7 CRC64;

Query Match 5.4%; Score 134; DB 1; Length 580;
Best Local Similarity 22.9%; Pred. No. 0.03;
Matches 75; Conservative 54; Mismatches 114; Indels 84; Gaps 18;

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 47.4586 Seconds  
(without alignments)  
3231.066 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHP1ASTETRRQAALGKLDP.....QETLDVAVARLFCBIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMBL 25:\*

2: sp\_archaea:\*

3: sp\_bacteria:\*

4: sp\_fungi:\*

5: sp\_human:\*

6: sp\_invertebrate:\*

7: sp\_mammal:\*

8: sp\_mmc:\*

9: sp\_organelle:\*

10: sp\_phase:\*

11: sp\_plant:\*

12: sp\_rodent:\*

13: sp\_virus:\*

14: sp\_vertebrate:\*

15: sp\_unclassified:\*

16: sp\_rvirus:\*

17: sp\_bacteriaph:\*

18: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	669.5	27.1	826	Q8VLS0	Q8VLS0 yersinia en
2	637	25.8	472	Q9LAE6	Q9LAE6 aeromonas h
3	379	15.4	546	Q8YQW1	Q8YQW1 anabaena sp
4	286	11.6	596	Q9RBJ7	Q9RBJ7 xanthomonas
5	246.5	10.0	600	P72289	P72289 rhizobium l
6	243.5	9.9	236	Q8YQW2	Q8YQW2 anabaena sp
7	179.5	7.3	428	Q56238	Q56238 thermus the
8	176.5	7.1	543	Q8FP71	Q8FP71 leptospira
9	176	7.1	368	Q37396	Q37396 chlorella v
10	175	7.1	848	Q8VVK9	Q8VVK9 corynebacte
11	174.5	7.1	546	Q9KVZ8	Q9KVZ8 xanthomonas
12	174.5	7.1	571	Q8VT54	Q8VT54 xanthomonas
13	172.5	7.0	862	Q9EXA5	Q9EXA5 neisseria m
14	165	6.7	487	Q25210	Q25210 helicobacte
15	163	6.6	540	Q53293	Q53293 acinetobact
16	161	6.5	543	Q9ZM08	Q9ZM08 helicobacte

17	159.5	6.5	435	2	Q93HR2	Q93hr2 thermus the
18	159	6.4	528	16	Q8PUU4	Q8puu4 xanthomonas
19	158.5	6.4	541	2	Q56752	Q56752 bergeyella
20	157	6.4	379	2	Q9KGT6	Q9kgt6 helicobacte
21	147.5	6.0	568	2	Q9KJH2	Q9kjh2 bacillus sp
22	147	6.0	379	16	Q25568	Q25568 helicobacte
23	147	6.0	579	2	Q9RKQ2	Q9rkq2 bacillus st
24	141	5.7	571	16	Q97DB4	Q97de4 clostridium
25	141	5.7	832	16	Q8UGB6	Q8ueb6 agrobacteri
26	140	5.7	555	16	Q9CCK4	Q9cck4 mycobacteri
27	137	5.5	514	16	Q9K001	Q9k001 neisseria m
28	136	5.5	684	16	Q8EGV1	Q8egy1 shewanella
29	135	5.5	553	16	P96868	P96868 mycobacteri
30	135	5.5	553	16	Q7TWV2	Q7twv2 mycobacteri
31	135	5.5	2162	2	Q9FB27	Q9fb27 streptomyce
32	134	5.4	514	16	Q9JVI6	Q9jvi6 neisseria m
33	134	5.4	534	16	Q8DUM7	Q8dum7 streptococc
34	134	5.4	800	2	Q93C65	Q93c65 thermus agu
35	134	5.4	1105	2	Q8GJ86	Q8gj86 thermus agu
36	133.5	5.4	381	16	Q9ZKT9	Q9zkt9 helicobacte
37	132	5.3	357	12	Q96719	Q96719 chlorella v
38	131.5	5.3	620	2	Q56788	Q56788 xanthomonas
39	131	5.3	1061	2	Q8RNV3	Q8rny3 hafnia alive
40	130.5	5.3	369	16	Q8PKD3	Q8pkd3 xanthomonas
41	130	5.3	526	16	Q92S33	Q92s33 rhizobium m
42	130	5.3	545	16	Q25223	Q25223 helicobacte
43	127.5	5.2	352	10	Q8L867	Q8l867 arabidopsis
44	127.5	5.2	380	10	Q9FK02	Q9fk02 arabidopsis
45	126.5	5.1	808	17	Q8PUN8	Q8pun8 methanosarc

## ALIGNMENTS

RESULT 1  
Q8VLS0 Q8VLS0 PRELIMINARY; PRT; 826 AA.  
AC Q8VLS0; 2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Methyltransferase-endonuclease.  
GN YENI  
OS Yersinia enterocolitica.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=IS1222;  
RA Rakin A.V.;  
RT "Characterization of the restriction-modification system of Y.  
RT enterocolitica biotype 1B.";  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ414030; CAC95150.1; -  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0004519; F:endonuclease activity; IEA.  
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.  
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . .; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006306; F:DNA methylation; IEA.  
DR InterPro; IPR002296; N12N6\_mtfase.  
DR InterPro; IPR002052; N6\_Mtase.  
DR PRINTS; PR00051; SAM bind.  
DR PROSITE; PS00092; N6\_Mtase; 1.  
KW Transferase; Endonuclease.  
SQ SEQUENCE 826 AA; 93348 MW; 148432FD8E798843 CRC64;

Query Match 27.1%; Score 669.5; DB 2; Length 826;  
Best Local Similarity 32.5%; Pred No. 2.1e-42;  
Matches 162; Conservative 107; Mismatches 186; Indels 43; Gaps 13;

QY	11	ROAALGKLDPTTQAVLGQFFTEPMKAAATLMASMLRVDDLRGTVRVLDPGAGVSLTAALVD	70
Db	9	RVKANANLVNKKGELQOFTSSSICIFMASLP--NELKGDISLDDPGCGPSLTAAPTE	66
QY	71	RLHATERPDVAHVVVAVETDPFVVPYLRATILEEC--RNAYGISVDLVE--GDYLL-----	120
Db	67	EVIRGARSLELHAIDIERKIKPFDLWLDKCVSASNAAGICKKIYPQINDYITAASVT	126
QY	121	--NOGAKLDGFDLVIANPPYGGKLASLSLARLATARAADVPNVYVAFVZAVISLKEQG	178
Db	127	KHDFGTEM--YTHCIINPPYKKTISADYRKLIISAIGIEANVLYAGVALAIWOLKQKG	183
QY	179	RGVFIYPRSWANGPYXQFRHMLTAVSLDILHVFESRTKFVADTKVQENVIYAFSVRP	238
Db	184	EMWAILPRSPCNGPYLPPFRNFIQHCALKHVHIFDSRSHAFSEDDVLQENIIHLVKNG	243
QY	239	QSSSVVL-----SRVANGHEESIASVPSFALSALVHDEDDDKIVHPA-----ESA	281
Db	244	IQESVKITSSPNSDFFPDQESNSVSASDMTV-RNIPFESLVNMLDKRFFHIAANRRDQS	302
QY	282	SVPSAARF--TLADLIGTVSTGKVVDPRNQYLTDLNLDASGVVPMVYQSNIRSGKIDWPO	339
Db	303	IIDLNVFYTSLNELGIVSVSTGPPVDFRLKSLDRENIE-PGAVPLIYVPHVL--NGVVDWPK	360
QY	340	VGARKPGCFVAVEDVALRQLLPQGSYVVVKQOTAKEDRRRVATAAWDGA--SRVALDNK	396
Db	361	-KSKKPNAINVSESRSRWLNQGVFVIVRPSKKEKRRIVATVYDGSPLGCEWIGFENK	419
QY	397	TNYLHESORPLEKNVARGIMLWNLSTVLDQYFRAFSGTQVNNAGILRLRPFICREDLILL	456
Db	420	LVNFHINKSGMDKOIAYGLSAFLNSMLLDKYVRLFGGHTQINATDLRSLHYDPDKSLQRI	479
QY	457	AKVVDPDGLPDQETLDVV	474
Db	480	GSYISSQGLSQENINEAI	497

```

RESULT 2
Q9LA86 PRELIMINARY; PRT; 472 AA.
ID
AC Q9LA86;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative modification methyltransferase (Fragment).
OS Aeromonas hydrophila.
OC Bacteriia; Proteobacteria; Gammaproteobacteria; Aeromonadales;
OC Aeromonadaceae; Aeromonas.
OX NCBI_TaxID=644;
RN [1]_TaxID=644;
RP SEQUENCE FROM N.A.
RC STRAIN=PPD134/91;
RA MEDLINE=20244644; PubMed=10784058;
RX Zhang Y.L., Ong C.T., Leung K.Y.;
RT "Molecular analysis of generic differences between virulent and
RT avirulent strains of Aeromonas hydrophila isolated from diseased
RT fish.";
RL Microbiology 146:999-1009 (2000).
DR EMBL; AF146609; AAF45040.2; -.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12NGMTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Methyltransferase; Transferase.
FT NON TER 472
SQ SEQUENCE 472 AA; 53084 MW; 24C8D9E6DF70B57B CRC64;
Query Match 25.8%; Score 637; DB 2; Length 472;
Query Local Similarity 34.2%; Pred. No. 2.7e-40;

```

[illegible]

```

RESULT 3
QBQYQW1 PRELIMINARY; PRT; 546 AA.
ID QBQYQW1;
AC AC QBQYQW1;
DT 01-MAR-2002 (TtEMBLrel. 20, Created)
DT 01-MAR-2002 (TtEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TtEMBLrel. 24, Last annotation update)
DT Type II site-specific deoxyribonuclease.
GN ALR3701.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CX NCBI_TaxID=103690;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=21595285; PubMed=11759840;
RX kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyano bacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003594; BAB75400.1; -.
DR PIR; AF2268; AF2268.
KW Complete proteome.
SQ SEQUENCE 546 AA; 62662 MW; BCA9A4C3CCB2A5E8 CRC64;

Query Match 15.4%; Score 379; DB 16; Length 546;
Best Local Similarity 38.3%; Pred.No.2.le-20;
Matches 88; Conservative 47; Mismatches 79; Indels 16; Gaps 6;

QY 266 VHEDDDKIVHFAESASVPSARE-----TLADLGIGVSTGKGVDFRNRQYLTDNLDA 318
DB 1 MNENDSDSYIIITN-SLEDSLRVQMDOFSSITDELGLAVSTGPVDFRLKSALRNYNE 59
QY 319 SGVVPMVYQSINRSKIDWPQGARKPKPGFVA--EDVALQLLI PQGSYVVVKRTAKD 376

```

```
Db 60 E-TVPELLYPAIKTKVLP--PKKPKRSIAIVQNETDKWLIPSGVYVLTFRSAKEE 115
QY 377 RRRVIAAVMD--GASRVADNKNYLVHESQRPLEKNVARGMLNLSNTVLDQYFRAFSGH 434
Db 116 KRRVVAACVPDAPVLGIENHLNYYHSGQGMNPDLAGLAAFLSSILLDSYFRFSGH 175
QY 435 TQVNAAGDLRLPFLCREDLILIAKVVPDGLPDQETLDAVAVARLFCPEIPES 484
Db 176 TQVNATDLRLIKYPCCKDDLKLGSGIGDSCLDQAOQLDTVWHKTLMSSEA 225

RESULT 4
Q9RBJ7
ID Q9RBJ7 PRELIMINARY; PRT; 596 AA.
AC Q9RBJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methyltransferase homolog M.Xphi.
GN XPHIM.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OC NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xcp 73;
RA Lai J.Y., Yang M.T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042157; AAP22366.1; -.
DR HSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12NGMTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase.
SQ SEQUENCE 596 AA; 64745 MW; 87AED38B1C2995F3 CRC64;

Query Match 11.6%; Score 286; DB 2; Length 596;
Best Local Similarity 26.3%; Pred. No. 3.2e-13;
Matches 132; Conservative 55; Mismatches 186; Indels 126; Gaps 20;

QY 27 GQFFTPMKAATLMASMLRVDDL-R-GTVRVLDPGAGVSLTAALVDRLHTE-----RDPVAV 81
Db 125 GVEYTPPALAERLDDQAAAGLQMSRAHVLDDPAAGAGFLVPAARLLKSLGDCSPAVAL 184
QY 82 HVVAV-----ETDPF-----VVPYL-----RATLEECRNAGVIGYDLVEGD 117
Db 185 HNLSARLGRFELDPFAAMWAQVFEAAALPLIVACGRPTAVLTVDG-----LSITKSG- 240
QY 118 YLLNQAKLDGPDLLVIANPPYKGLASDSLARIATTARAVDVPNVVAVFWRVAVISLKEQ 177
Db 241 -----FDLVGNPPFGRLKLAERREYFSGSLYGHANLYGLFMDLAVLAKPD 288
QY 178 GRGVFIVPSWANGPYVROFR---HMLMTAVSLDILHVFESRTKVPFADTKVQENIVAF 234
Db 289 GLVSFLTPSSFLAGFYKKNLRAVLHKEAPPVSLDFV---TARKGVFDD---VLQETVLATY 343
QY 235 S-----VRPQSSSVLRSVAHGEESIASVPFSALVHDEDDDKIVHFEASAS 282
Db 344 RXGAKRARAVVSFIEAQPGVPVKAEPAGHTLPKRTATAPWFLPRHADE-----AELAK 396
QY 283 VPSAARFTLADLIGIVSTGKVVDNRQYLTNDLSDAGSVVPMYQSNIRS-GKI----- 335
Db 397 RLRAKSAPLADMGYKVTGTPLVNRRKPOLCDSEEA-GTVPLVWAEVSTSDGFFVLRAEK 455
QY 336 -----DWQVGARKPGQGVAVEDVALROLQLPQSGSVVVKVQTAKEDRRRVIA 382
Db 456 RNHKPFLRLQPGDDMLV--RKP-----CVLIQRTTAKEARRLIA 494
```

```
QY 383 AWDGA-----SRVALDNKNYLVHESQRPLEKNVARG---LMLWLNSTVLDQYFRAFSGH 434
Db 495 AEMPAKFIRKHAGVTIENHLNM-----IPTVENPAVSPALLAAFLNSDAADRAFCMSG 550
QY 435 TQVNAAGDLRLPFLCREDL 453
Db 551 VAVSAVELENIPLPTASDL 569

RESULT 5
P72289
ID P72289 PRELIMINARY; PRT; 600 AA.
AC P72289;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methylase.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OC NCBI_TaxID=384;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98055154; PubMed=9393436;
RA Rochepeau P., Selinger L.B., Hynes M.F.;
RT "Transposon-like structure of a new plasmid encoded restriction-
modification system in Rhizobium leguminosarum VF39SM.";
RL Mol. Gen. Genet. 256:387-396(1997).
DR EMBL; X99520; CAA67873.1; -.
DR HSP; P14385; 2ADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR00507; N12NGMTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 600 AA; 65707 MW; C3909584B71D69DF CRC64;

Query Match 10.0%; Score 246.5; DB 2; Length 600;
Best Local Similarity 25.8%; Pred. No. 3.4e-10;
Matches 137; Conservative 70; Mismatches 230; Indels 95; Gaps 25;

QY 3 PIASTRTROAALGKLDL-TTQAVLGQFFTPMKAATLMASMLRVDDL-RG-----TVRVLD 57
Db 104 PIESAFLQSACVTAMLPQDIRSEWGAFTY-----PALTKLMELAQEAGIDWRAARVLD 160
QY 58 GAGVSLTAALVDRLHT-----ERPD-VAVVAVVAVETDPFVVPYLRATLEECRNAY 108
Db 161 ACGGAGFLLFVALRMQOALQALSPGELLDFHAGLSGFDIDPFAAQLTQTWLE---IAFA 217
QY 109 ISVDLVEG---DYLLNQAKLDG-----FDLVIANPPYKGLASDSLARIATTARAVDP 160
Db 218 -SLSMOTGRPFPAFVTRVNCNSLEQPVSSKEFDLVIGNPPYGRVLRNARLRERYRSLFG 276
QY 161 NVYVAFVWRVAVISLKEQGRGVFIVPSWANGPYVROFRHMLMTAVSLDILHVFESRTKVP 220
Db 277 NMVGLFTDLALQWARKGVAVVPTGFLAGFYFKALRALAKADAPPFAIDITERRGIF 336
QY 221 AOTKVQENIVAF-----SVRPQSSSVLRSVAHGEES-----IASVSPFS 263
Db 337 DD--VLQEAALLATYRRGGSLGSPTVHYFSVNGTAQVTHAGEFHLPKDASQPLARVP-- 392
QY 264 ALVHDEDDDKIVHFEASASVPSAARFTLADLIGIVSTGKV-----DFNRQYLTNDL 317
Db 393 -----DDGVLY--KQLSRLPS-----RLKDWGKYSTGTPLVWNRKQDPGRGA----- 433
QY 318 ASGVVPMYQSNIRS-GKIDW-POVGARKPGQGVAVEDVALROLQLPQSGSVVVKVQTAK 375
Db 434 GKDTFFVIAEAVSADGKFSFAEKRNQHPFKVREPQOWLEVSVP---CVLIQRTTSKE 490
```

```
QY 376 DRRVIAAVWDG-----SRVALDNKTNLYHESQ-RLEKNVARGMLMLNSTVLDOYFR 429
Db 491 QARLIAAELEPEAFIKAHGRVIVENHLNMVPTAKGPRVSTAVVAV--LNSKIADRAFR 548
QY 430 AFSGHTOVNAGDLRLPLFLCKREDILILLAKVVPDGL--PDQETILDVAVARLF 478
Db 549 CIGSVAVSAFELEALPLPKPEAL---KAVGDLILAKHADQAVIDAALERLF 596

RESULT 6
Q8YQW2 PRELIMINARY; PRT; 236 AA.
AC Q8YQW2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Site-specific DNA-methyltransferase.
GN ALB3700.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213 (2001).
DR EMBL; AF003594; BAB75399.1; -.
DR PIR; A82268; AE2268.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR PRINTS; PR00507; N12N6MTFRASE.
DR TRANSFERASE; Methyltransferase; Complete proteome.
KW TRANSFERASE; Methyltransferase; Complete proteome.
SQ SEQUENCE 236 AA; 26734 MW; 63C1EC76516FF9ED CRC64;

Query Match 9.9%; Score 243.5; DB 16; Length 236;
Best Local Similarity 31.3%; Pred. No. 1.4e-10; Indels 17; Gaps 6;
Matches 68; Conservative 41; Mismatches 91;

QY 6 STETTRQAALGKLDPTQAVLGQFFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSILT 65
Db 8 STDINRINFSSRLNLRHRELGOFTLPAPLRFVVG--QFNLSLGHISLIDPCAGVGSILT 65
QY 66 AALVDRLHTERPDV-AHVAVVETDPFVVPYRATLEECRNAY---GI---SYDLVEGDYL 119
Db 66 AAFVERLLANSHEVKSCFTIYVEEPIFLPKQCLIDCCTALENGIAANYCLYEKNF- 124
QY 120 LNQAKLDGP-----PDLVIANPPYKGLASDSLRLATTARAVDVNPNVYVAVWRAV 171
Db 125 IDANSINPLSTSAINTFHALINPLYKKNKNSIEKKIISQMGITVNLISAFWLII 184
QY 172 ISLKEQGRGVFIIPRSWANGPYRQFRHMLMTAVSLD 208
Db 185 IRLVEGEIVAITPRSCFNGSYRHRFRKSFLEQMKLE 221

RESULT 7
Q56238 PRELIMINARY; PRT; 428 AA.
AC Q56238;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Methylase.

GN M. TTHB81.
OS Thermus thermophilus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HB8;
RX MEDLINE=92201699; PubMed=1339363;
RA Brany F., Danzitz M., Zebala J., Mayer A.;
RT "Cloning and sequencing of genes encoding the TthB81 restriction and
RT modification enzymes: comparison with the isoschizomeric TaqI
RT enzymes.";
RL Gene 112:31-95 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=HB8;
RX MEDLINE=92201707; PubMed=1551602;
RA Barany F., Siatko B., Danzitz M., Cowburn D., Schildkraut I.,
RA Wilson G.G.;
RT "The corrected nucleotide sequences of the TaqI restriction and
RT modification enzymes reveal a thirteen-codon overlap.";
RL Gene 112:91-95 (1992).
DR EMBL; M76882; AAA27491.1; -.
DR HSSP; P14385; AADM.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. .; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000051; SAM_bind.
DR PRINTS; PR00507; N12N6MTFRASE.
DR PROSITE; PS00092; N6_MTASE; 1.
SQ SEQUENCE 428 AA; 48478 MW; F24945880651D9CD CRC64;

Query Match 7.3%; Score 179.5; DB 2; Length 428;
Best Local Similarity 22.4%; Pred. No. 2.8e-05; Indels 153; Gaps 22;
Matches 110; Conservative 60; Mismatches 168;

QY 18 LDPTTQA-VLGOFFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSILTALVDRLHTER 76
Db 7 LDPTASGRSLGRVETP---PGLVRFMVGLAEARKGVRLPEACADGPFL-----RAFREA 58
QY 77 PDVAVHVAVETDPFVW---PYLRATLEECRNAYGISYDLVEGDYLLNQAKLDGPDFLY 133
Db 59 HGTGYRFVGVGEIDPHALDLPWAEG-----VVADFLMEFGE---AFDLI 100
QY 134 IANPPYKGLASDSLRL-----ATTARAVDPNPNVYVAVWRAVISLKEQGRGVF 182
Db 101 LGNPPYGIYGEASKYPIHVLREVKGLYKKTLSWKKGKYNLYGAFIEKSVRLREGTLVF 160
QY 183 IYPRSWANGPYRQFRHMLMTAVSLDILH---VFESRTKVA-----DTK 224
Db 161 VVPATWLVLDDPSLLRSFLAREGRIEYVYLGVEFPGR-KVSAVLFKRGKGLALWDTR 219
QY 225 VKQENVI-VAFSVRPOSSSVIISRSVAHGEESTASSVPFSALVHDEDDDKIVHFAFSASV 283
Db 220 RDGETTPLLWSEKPEWKGEIIRFETGWTREMEASGPPLGSLFH-----I 264
QY 284 PSNAARFTLADLIGVSTGKGVDPFRNQYLTDNLDSGVVPMVYQSNIRSGKIDW----- 337
Db 265 PFAARSP-----EFKHPAVOKEPE-FGLVPLVTGRNLKPGWIDYESNHSG 309
QY 338 ---POVGARKPGQFVAVEDVALRQLLPQGSYVYVVKRQTAKEDRRRRVIAAVWDGAS----- 389
Db 310 LMPKERAKELRDFYATPHL-----VVAHTKGTG-----VVAWDERAYPWRE 352
QY 390 -----RVALDNKTNLYHESORPLEKNVARGMLMLNSTVLDOY-----FRFSGHTQV 437
Db 353 EFHLLPKGVELD-----PL-----FLVWLNDSKIQEVYKTYLRDFVPHLTL 395
QY 438 NAGDLRLPL 448
```







RT pylori."  
RL Nature 388:539-547(1997).  
DR EMBL; AE000561; AAD07525.1; -.  
DR PIR; G64577; G64577.  
DR TIGR; HP0463; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.  
DR GO; GO:0009007; F:site-specific DNA-methyltransferase (adenin. . .; IEA.  
DR GO; GO:0006306; P:DNA methylation; IEA.  
DR GO; GO:0006304; P:DNA modification; IEA.  
DR InterPro; IPR003665; Methylase M.  
DR InterPro; IPR002296; N12N6 mtfrase.  
DR InterPro; IPR003356; N6 DNA\_Mtase.  
DR InterPro; IPR002052; N6 Mtase.  
DR InterPro; IPR000051; SAM bind.  
DR Pfam; PF02384; N6 Mtase; 1.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
DR Hypothetical protein; Complete proteome.  
KW HYPOTHETICAL PROTEIN; Complete proteome.  
SQ SEQUENCE 487 AA; 55800 MW; 955DFF0EED48867D CRC64;  
  
Query Match 6.7%; Score 165; DB 16; Length 487;  
Best Local Similarity 24.4%; Pred. No. 0.00044;  
Matches 63; Conservative 49; Mismatches 90; Indels 56; Gaps 12;  
  
QY 28 QFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSLSLAALVDRHLHTERPDVAVHVAVE 87  
DB 208 EYTPLSIASIIAKLL-INETRNKVIYDPSAGTGTLLMALAHQIGTD---SCTLYAQD 262  
  
QY 88 TDPFVVPVLRATL-----EECRNAYGISYDLVEGDYLLNQAKLD--GPFDLVIANPPY 139  
DB 263 ISQSLRMKLNLIINDLTHSUNA-----IEGNTLTNPYHSDPKGMDYIVSNPPF 315  
  
QY 140 GKLASDSLRLATTAR--AVDVPN-----VYVAFWVRRAVISLKEQGRGVIVPRS 187  
DB 316 KLDFSENEHAEISQNKDFFLGVPNPKNKSQMPIYTLFFQHLNMLSNKGKGAIVPTG 375  
  
QY 188 WANGPYRQFRHMTAVSLDIL-HVFESR-----TKVFADTKVQENV-IVAFSV 236  
DB 376 FISA-----KSGVENKIIRHLVDERLVYGVCMPSQVFANTGT---NVSIFFOK 422  
  
QY 237 RQSSSVLSRSVAHCEE 254  
DB 423 TESAKEWVLIDASKLGE 440  
  
DR InterPro; IPR002052; N6\_Mtase.  
DR InterPro; IPR000051; SAM bind.  
DR PRINTS; PR00507; N12N6MTFRASE.  
DR PROSITE; PS00092; N6\_MTASE; 1.  
DR Hypothetical protein; Complete proteome.  
KW HYPOTHETICAL PROTEIN; Complete proteome.  
SQ SEQUENCE 540 AA; 63078 MW; 79CD2BF714352CA5 CRC64;  
  
Query Match 6.6%; Score 163; DB 2; Length 540;  
Best Local Similarity 18.7%; Pred. No. 0.00074;  
Matches 95; Conservative 84; Mismatches 176; Indels 154; Gaps 19;  
  
QY 28 QFTPMKAATLMASMLRVDDLRGTVRVLDPCAGVGSLSLAALVDRHLHTERPDVAVHVAVE 87  
DB 31 QFTTFFPIAYAMAKWILGNKQLKT--VLEPAFGLVGFSRAILSQ-----QKEINIKGFE 82  
  
QY 88 TDPFVVPVLRATLLEECRNAYGISYDLVEGDYLLNQGA-KLDGPPFDLVIANPPYGKL---- 142  
DB 83 VDETIFENAKYFDDFENV-----NILLQDYNDKMKYDG---IICNPPYKFKHDYD 133  
  
QY 143 ASDSLARLATTARAV--DVPNVYVAFWVRRAVISLKEQGRGVIVPRSWANGPYRQFRHW 200  
DB 134 KKNILKEIETNLKCKLNGFTNLYTLFLKLSIHQLSQNGRCAYIIPSEFLNSDYGKLVKTY 193  
  
QY 201 LMTAVSLDILHVFESRKYVFADTKVQENVIVAFSVRQSSSVLSRSVAHGEESIASV 260  
DB 194 LIKSKTLRHIIIVDFEENVFDDALT-----TASIILC-----ANDNITDKV 234  
  
QY 261 PFSALVH-----DEDDDKIVHFAESASV-----PSAARF----- 289  
DB 235 QFNIIQSLODLSKDEIINKYPNFELETQTYNFSDLNPEIKWKAYYQKNSIKFKPNLVPF 294  
  
QY 290 -TLADLIGIGVSTGKVVDPRNQYLTDLN-----DASGVVPMV----- 325  
DB 295 STYAKVVRGIATG-----SNEYFTFNLKSAKEFNIDQYLLPCICSAKDAKTSFTTKOD 348  
  
QY 326 -----YQSNTRSKID-----WPQVGARKP----- 345  
DB 349 FEELKKSKSVFLFNAQNSTDKNISSYQKGESEINKRFLTASRTPWVSLNRKAPW 408  
  
QY 346 -----QGFVAVEDVALRQLLPQGSYVVVVKQTAKEDRRRVIAA--VMDGASRVALDNKT 397  
DB 409 VSVFNRSGLRFRIRNEANISNLTSYHICIIQNKQVSEIDIDLLFAVLLTDTAKQIFEDNSR 468  
  
QY 398 NYLHESQRPLEKNVARGMLMLNLTVDQ 426  
DB 469 QYXGNGIQKFEPNDLNGKMWLDLG--LLDK 495  
  
Search completed: October 1, 2004, 16:10:27  
Job time : 51.4586 secs

RESULT 15  
Q53293 ID Q53293 PRELIMINARY; PRT; 540 AA.  
AC Q53293;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Methyltransferase  
OS Acinetobacter calcoaceticus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Moraxellaceae; Acinetobacter.  
OX NCBI\_TaxID=471;  
RN [1]\_SEQUENCE FROM N.A.  
RP MEDLINE=91345839; PubMed=1368703;  
RX Kawakami B., Christophe H., Nagatomo M., Oka M.;  
RT "Cloning and nucleotide sequences of the AccI restriction-modification  
RT genes in Acinetobacter calcoaceticus."  
RL Agric. Biol. Chem. 55:1553-1559(1991).  
DR EMBL; S50843; AAC60387.1; -.  
DR GO; GO:0003677; F:DNA binding; IEA.  
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.  
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransferase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006306; P:DNA methylation; IEA.  
DR InterPro; IPR002296; N12N6\_mtfrase.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 63.6786 Seconds  
(without alignments)  
2156.424 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHPASTETRRQAALGKLDP.....QETLDVAVVRLFCBIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Geneseqp29Jan04.\*  
2: Geneseqp1980s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	223	9.0	1068	6	ABR43481	Abm67143 Photorhab
2	174	7.0	550	6	ABM67143	Abm67143 Photorhab
3	164	6.6	543	2	AAW20785	Abp79018 N. gonorr
4	151	6.1	533	6	ABP79018	Abp79018 N. gonorr
5	147	6.0	626	3	AAV98485	Aay88485 Draill me
6	141	5.7	571	6	ABU23716	Abu3716 Protein e
7	137	5.5	514	3	AAV75218	Aay75218 Neisseria
8	137	5.5	519	6	ABM70049	Abm70049 Photorhab
9	135	5.5	510	6	ABP78365	Abp78365 N. gonorr
10	135	5.5	514	3	AAV75217	Aay75217 Neisseria
11	135	5.5	514	3	AAU37303	Abu37303 Protein e
12	135	5.5	2162	3	AAU07560	Aay07560 Protein e
13	134	5.4	514	3	AAV75219	Aay75219 Neisseria
14	134	5.4	514	6	ABU37930	Abu37930 Protein e
15	134	5.4	534	6	ABU44389	Abu44389 Protein e
16	130	5.3	545	2	AAW98867	Aaw98867 H. pylori
17	126	5.1	502	2	AAU15226	Aau15226 HincII re
18	125.5	5.1	317	4	AAU81134	Aag81134 Mycobacte
19	125.5	5.1	317	6	ABU34416	Abu34416 Protein e
20	125.5	5.1	317	6	ABU36560	Abu36560 Protein e
21	124.5	5.0	295	6	ABU25782	Abu25782 Protein e
22	124.5	5.0	511	4	AAU35144	Aau35144 Enterococ
23	124.5	5.0	530	6	ABU28930	Abu28930 Protein e
24	122	4.9	358	2	AAW98649	Aaw98649 H. pylori
25	122	4.9	530	6	ABU29993	Abu29993 Protein e

## ALIGNMENTS

RESULT 1

ABR43481  
ID ABR43481 standard; protein; 1068 AA.

XX AC ABR43481;

XX 23-OCT-2003 (revised)

DT 16-JUL-2003 (first entry)

XX DE Bacillus stearothermophilus B61 methylase protein SEQ ID NO:11.

XX KW Bacillus stearothermophilus B61; BsmI; methylase; bsmBIM; enzyme; molecular cloning; gene characterisation.

XX OS Geobacillus stearothermophilus.

XX PN EP1298212-A2.

XX PD 02-APR-2003.

XX PF 27-SEP-2002; 2002EP-00256756.

XX PR 28-SEP-2001; 2001US-00966997.

XX PA (NEWE ) NEW ENGLAND BIOLABS INC.

XX PI Xu S, Dore A, Hume A, Pellitier J, Zhou J;

XX DR WPI; 2003-395598/38.

XX N-PSDB; ACC69532.

XX PT New DNA segment coding for the BsmBI restriction endonuclease and/or BsmBI methylase, useful for molecular cloning and gene characterization, and in producing restriction endonucleases and modification methylases.

XX PS Example 1; Fig 2A-D; 35pp; English.

XX CC The present sequence represents Bacillus stearothermophilus B61 BsmBI methylase (bsmBIM). The present invention also describes BsmBI restriction endonuclease (bsmBIR) isolated from B. stearothermophilus B61. BsmBIR and bsmBIM can be isolated from a DNA segment obtainable from American Type Culture Collection (ATCC) Accession Number PTA-3739. The DNA segment is useful in molecular cloning and gene characterisation, and in producing BsmBI restriction endonucleases and modification methylases. (Updated on 23-OCT-2003 to standardise OS field)

XX SQ Sequence 1068 AA;

Adc96965 E. faeciu  
Aay04661 L.lactis  
Abp25613 Streptoco  
Adc96966 E. faeciu  
Abu48789 Protein e  
Aaw98507 H. pylori  
Abu02489 S. pneumo  
Abu44324 Protein e  
Abb54003 Lactococc  
Abu16328 Protein e  
Aay81754 Streptoco  
Aau38061 Streptoco  
Abu02552 S. pneumo  
Abu46269 Protein e  
Abu20637 Protein e  
Abu30617 Protein e  
Abu16331 Protein e  
Abm71646 Staphyloc  
Aau37589 Streptoco  
Abu46253 Protein e

Query Match 9.0%; Score 223; DB 6; Length 1068;  
 Best Local Similarity 20.7%; Pred. No. 3.8e-13;  
 Matches 112; Conservative 86; Mismatches 219; Indels 124; Gaps 21;

QY 21 TTQAVLGQFTPMKAAATMASMLRVDLGRGTVRVLDPCAGVGSUTA-ALVDRHLHTERPDV 79  
 Db 118 TASKAYGYFTPIISLGTWVKALADKPKNLKSIVDPACGIGSLALALI-----YNPEI 172

QY 80 AVHVAVVETDPFVVPYIATLEECRNAYGIS--VDLVEG---DYLLNGQAKLDGPFDLVI 134  
 Db 173 E-NVVGIELDSFTANISHKLLVRISKDLGITPKIKIINQNFYLVLYNVEEHEKPFLLI 231

QY 135 ANPYGKLI-----ASDLARLATTARAVDVP 160  
 Db 232 MNPPYGRVRLKSLTNKETSGLTEGISELEKLRBEETILNADLRKXKFSVGLGKTP 291

QY 161 NVYVAFWVRVVISLKEQGRGVFIVPRSWANGPYVROFRHMLTAVSLDILHVFESRTKVF 220  
 Db 292 EYSKVFLAISTKIVKQNGYVIAITPSSWLGDESGRELKYLVENHGISCIWNFKESAKLF 351

QY 221 ADTKVQBNVIAFVRPQSSVLSRVAHGEESIASVPSALVHDEDDDKIVHFAES 280  
 Db 352 SG--VNQPTTVKIKVNSKIEIQG-----PLSSL---EELGRDIQYLDI 393

QY 281 ASVP--SAARFLADLGIG-----VSTGKVDVFRNQYLTNDLSDAGVWPM 324  
 Db 394 CNIKYSPWYRPOCGNERAKILSKLNHAPLSHKYINLRGELDTSHKOLLSDNPN 453

QY 325 VQOSNRSKIDWPVGARKPO-----GFVAVEDVAL-----ROLIQQ 362  
 Db 454 HWRL-IRGDHVE--KENLKNPSESEKLGIV--DHQLFKRMGKSNKLRHKIKNWRTILPQ 507

QY 363 GSYVYVVKQTAKEDERRRVIAAVWDGASRVALDNKNYLYH-ESQRPLEKXVARGIWLNS 421  
 Db 508 CSYM-----NKKRIEACIVEPNNIIA--NSCNYITLEDNCNELVDNLLL-LCAIINS 556

QY 422 TVLDQYFRAFSGHTQVWAGDLRLPLC---REDLILAKVVPDGLPQDTLDVAVARL 477  
 Db 557 AVIEWREFLNSNNHVSNEYDEFPFKPDTETEMTLKGFHFKPIENWKSIEALIALM 616

QY 478 F 478  
 Db 617 Y 617

RESULT 2  
 ID ABM67143  
 XX ABM67143 standard; protein; 550 AA.  
 AC ABM67143;  
 XX 20-NOV-2003 (first entry)  
 DE Photorhabdus luminescens protein sequence #240.  
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 XX whooping cough.  
 OS Photorhabdus luminescens.  
 XX WO200294867-A2.  
 XX 28-NOV-2002.  
 XX 07-FEB-2002; 2002WO-IB003040.  
 PF 07-FEB-2001; 2001FR-00001659.  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX

PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 DR WPI; 2003-148459/14.  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 PT Claim 2; SEQ ID NO 240; 1205pp; French.  
 XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX

Query Match 7.0%; Score 174; DB 6; Length 550;  
 Best Local Similarity 21.7%; Pred. No. 1.6e-08;  
 Matches 106; Conservative 71; Mismatches 176; Indels 136; Gaps 23;

QY 53 RVLDPCAGVGSUTAALVDRL-----HTERPD-----VAHVAVETDPF-----VVP 94  
 Db 39 RLLEPSFGGDFLLPIERLLSAWRAARPSGALDELDTAIRAVELHHTFKSTKAAVYT 98

QY 95 YLRATLEECRNAYGI-SYDLVEGDYLLNQAKLDGPFDLVIANPPY-----GKLASDS 146  
 Db 99 LLRRQCIANAARVTLASYWLSQDFLL---APLEGQDFVGNPPYVROELIPAPLLAEY 155

QY 147 LARLATTARADVVPNVYVAFVVRVAVISLKEQGRGVFIVPRSWANGPYVROFRHMLTAVS 206  
 Db 156 RSKYLT---MYDRADIYIFFIERSLSVSDGSLGFCADRWMKNRYGGPLR-----SLV 207

QY 207 LDILHVFESRTKVF---ADTKVKQENVI-----VAFSVRQSSSVVLS-- 246  
 Db 208 AERFHL-----KVYVDMVDTPAFHSVDIAYPAITISREVPRTRIAYRPAIDPATLTAL 262

QY 247 -----RSVAHGEESIASVPSALVHDEDDDKIVHFAESASVPSAA 287  
 Db 263 AGLKAPSLPKVGPARELAQVTNGSE-----PW--LJESFQMAVIRLE-----G 307

QY 288 RFTLAD-----LGIGVSTGKVDVFRNRYQVLT---NLDASGVVPMYQSNIRSGKIDWP 338  
 Db 308 DFFLLEAGCKVIGVATG-----ADKAFIGDFDALNVEPDVKVPLVTTKIDMTGEVWR 362

QY 339 QVGARKP---QGFVAVEDV-ALRQLLPQSGSVVVRQTKAKE-----D 376  
 Db 363 GLGVINPFABQGGVLVDLDEYPRLLRYLEARREVIAGRHCAQKAPANNWYTRIDITPSLAA 422

QY 377 RRVIAAVWDGASRVALDNKNYLYHESQRPLEKN--VARGMLMLNSTVLDQYFRAFSGH 434  
 Db 423 RPKLLIPDKQAHIVYENGELYPHNNLYYVTSDDDLRALQAVLLSAVTRLFVATYS-- 480

QY 435 TVNAGDLR 443  
 Db 481 TKMRGGFLR 489

RESULT 3  
 AAW20785 ID AAW20785 standard; protein; 543 AA.  
 XX AC AAW20785;  
 XX DT 16-JUL-1997 (first entry)  
 XX DE H. pylori cytoplasmic protein, 07gp11807orf35.  
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacterium; life cycle; activator;  
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;  
 KW diagnosis.  
 XX OS Helicobacter pylori.  
 XX PN WO9640893-A1.  
 XX PD 19-DEC-1996.  
 XX PF 06-JUN-1996; 96WO-US009122.  
 XX PR 07-JUN-1995; 95US-00487032.  
 XX PR 01-APR-1996; 96US-00630405.  
 XX PA (ASTR ) ASTRA AB.  
 XX PI Smith D, Berglinth OT, Mellgaerd BL;  
 XX DR WPI; 1997-052306/05.  
 XX DR N-PSDB; AAT68038.  
 XX PT Helicobacter pylori nucleic acid sequences and related polypeptide(s) -  
 PT useful for vaccines to treat or prevent H. pylori infection, and to  
 PT detect Helicobacter.  
 XX PS Claim 61; Page 1194-95; 1481pp; English.  
 XX CC This sequence represents a H. pylori cytoplasmic protein involved in  
 CC genome replication, transcription, recombination and repair. The protein  
 CC may be used in a vaccine to prevent or treat H. pylori infection or to  
 CC identify H. pylori polypeptide binding compounds, useful as potential H.  
 CC pylori life cycle activators or inhibitors. The genomic sequence of H.  
 CC pylori (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were analysed for  
 CC ORF of at least 180 nucleotides, and the predicted coding regions defined  
 CC by computer evaluation. To identify likely H. pylori antigens for vaccine  
 CC development, the amino acid sequences predicted from various ORF were  
 CC analysed for significant homology to other known or exported membrane  
 CC proteins. Having identified and determined the sequences of interest,  
 CC particular regions can be isolated from H. pylori by PCR amplification  
 CC for recombinant polypeptide production, e.g. in E. coli hosts  
 XX SQ Sequence 543 AA;  
 Query Match 6.6%; Score 164; DB 2; Length 543;  
 Best Local Similarity 23.3%; Pred.No.1.8e-07;  
 Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;  
 QY 28 QFTPMKAATLMSMLRVLDLGRVRLDPGAGVSLTAALVDRHLTERPDVAHVAVVE 87  
 Db 208 EYITPLSIASIIAKLL-VNKPTQSVKIYDPSAGTGLLMAHAHQIGTD--SCTLYAQDIS 264  
 QY 88 TDFVFPVPLRATLECRNAYGTSYDL---VEGDYLLNQAKLD--GPFDLVIANPFPYGLK 142  
 Db 265 QKSLRMLKLNLIIND-----LTHSLRHAIEGNTLNPVHSDKHGKMDFFVSNPPFKLD 318  
 QY 143 ASDSLARLATTAR--AVDYPN-----VYVAFWRAVISLKEQGQGVFIVR---S 187  
 Db 319 FSNHAEISQNKNDPFLGVFNIPKNDKSMPTITLFFQCLNMLSPKKGAIIVTGTFTS 378

QY 188 WANGPYRQFRLMLTAYSLDLHLVFSRTKVADTKVKQENV-IVAFSVRQSSSVVLS 246  
 Db 379 AKSGVNNKVRHLVDERL---VYGVICMPQSVFANTGT---NVSIIFFQKTPSAKEVILI 432  
 QY 247 RSVAGGEE-----SIASSVPFSGALV-HDEDDDDKIVHFAESA 281  
 Db 433 DASKLGEYTENKKNKTLRPSDMDLILLETQNKAPKSDFCALVSFDEITEK-----NY 486  
 QY 282 SVPSAARTLADLIGVSTGKVDPRN--ROYLTD 314  
 Db 487 SLNPGQVFTIEDTSETISQ---ABFENLMQOYSSE 518  
 RESULT 4  
 ABP79018 ID ABP79018 standard; protein; 533 AA.  
 XX AC ABP79018;  
 XX DT 07-MAR-2003 (first entry)  
 XX DE N. gonorrhoeae amino acid sequence SEQ ID 4566.  
 XX KW Antibacterial; infection; vaccine; gene therapy.  
 XX OS Neisseria gonorrhoeae.  
 XX PN WO200279243-A2.  
 XX PD 10-OCT-2002.  
 XX PF 12-FEB-2002; 2002WO-IB002069.  
 XX PR 12-FEB-2001; 2001GB-00003424.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PI Fontana MR, Pizza M, Masignani V, Monaci E;  
 XX DR WPI; 2003-058415/05.  
 XX DR N-PSDB; ABZ39988.  
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
 PT medicament for treating or preventing N. gonorrhoeae infection.  
 XX PS Disclosure; Page 523; 815pp; English.  
 XX CC The present invention relates to proteins from Neisseria gonorrhoeae.  
 CC Also disclosed are the nucleic acid molecules encoding the proteins and  
 CC antibodies that specifically bind to the proteins. The composition  
 CC comprising the protein, nucleic acid or antibody is useful for the  
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
 CC infection, this may be in the form of a vaccine or gene therapy.  
 CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
 CC molecules of the invention  
 XX SQ Sequence 533 AA;  
 Query Match 6.1%; Score 151; DB 6; Length 533;  
 Best Local Similarity 21.5%; Pred.No.4e-06;  
 Matches 79; Conservative 65; Mismatches 128; Indels 96; Gaps 17;  
 QY 28 QFTPMKAATLMSMLRVLDLGRVRLDPGAGVSLTAALVDRHLTERPDVAHVAVVE 84  
 Db 190 EYITPLSIAIIMADILVPEVREGVRSVDVDPGAGSGLT---LMNVAH----- 235  
 QY 85 AVETDFVFPVPLRATLECRNAYGTSY-----DLVEGDYLLNQAK-LDG---PF 130  
 Db 236 AIGEDKCM-I-YTQDISQKSSNLLRLNLIINLVHSLNNVQGNITLSPAHKDASGLKKF 294  
 QY 131 DLVIANPYPY-----CKLASD-----SLARLATTARAVDVPNVYVAFWRAVISL 174

Db 295 DFIVSNPPFKLDRSDFRDRLESDENHERFFAGIPKIPKTK--EKMEIYQLFIOHILFSL 352  
Qy 175 KEQGRGVFIIPRSW--ANGPYRQFRHMLTAVSLDILHVPESRTKVPADTKVKQENVIV 232  
Db 353 KENGKAAVLPTGPIITAKSGIDKIREYLVENKML--AGVVMSPSNIPATTTGT--NWSI 407  
Qy 233 AFSVRPOSSVVLRSVAHGEESTASSVPFSALVHDEDDDKI VHPAESASVPSAARFTLA 292  
Db 408 LFDKTKNRKQVLLDASGLGK--IKDGNQKTVLUSCEEQKICN----- 450  
Qy 293 DLGIGVSTGKVVDFERNROYLDNLNDASGVWPMVYQSNIRSG-----KIDWPQVG---A 342  
Db 451 -----TTNKQAVEFSVVVGYDEIKAKNHSLSAGQFEVKIDYVDISADBEFA 498  
Qy 343 RKQGFVA 350  
Db 499 QKIAGFSA 506

RESULT 5  
ID AAY88485 standard; protein; 626 AA.

XX AC AAY88485;  
XX DT 07-AUG-2000 (first entry)  
XX DE DraIII methylase DraIIIM amino acid sequence.  
XX KW DraIII; restriction endonuclease; DraIIIM; two plasmid system; pHKU5;  
XX KW PHK77; methylase.  
XX OS Deinococcus radiophilus.

FX Key Location/Qualifiers  
FT Misc-difference 399 /label= Unknown  
FT /note= "Encoded by NAT"  
FT Misc-difference 521 /label= Unknown  
FT /note= "Encoded by NTC"

XX PN US6048719-A.  
XX PD 11-APR-2000.  
XX PF 22-JAN-1999; 99US-00235246.  
XX PR 22-JAN-1999; 99US-00235246.  
XX PA (NEW ) NEW ENGLAND BIOLABS INC.  
XX PI Kong H, Dalton MA, Higgins LS;  
XX DR WPI; 2000-316897/27.  
XX DR N-PSDB; AAL15710.

XX PT Cloning and expressing an endonuclease gene by comprising two plasmid system.  
XX PS Disclosure; Fig 4; 25pp; English.  
XX CC This sequence represents the Deinococcus radiophilus (DraIII) restriction modification system methylase DraIIIM amino acid sequence. The DraIII endonuclease is a type II restriction endonuclease that recognises the DNA sequence 5'CACNNGTG 3', and cleaves the phosphodiester bond 5' to the first G of the recognition sequence to produce a 3 base 3' extension. The invention relates to the DraIII endonuclease gene (DraIIIR) and a method for its cloning and expression. The method comprises the use of a two plasmid system, using plasmids pHKU5 and pHK77. The invention also comprises a vector containing the DNA isolated by the method, a host cell transformed by the vector, and a method for culturing the host cell and producing a DraIII restriction endonuclease

XX SQ Sequence 626 AA;  
Query Match 6.0%; Score 147; DB 3; Length 626;  
Best Local Similarity 19.0%; Pred. No. 1.4e-05;  
Matches 89; Conservative 74; Mismatches 166; Indels 120; Gaps 18;  
Qy 12 QAALGKLDPTTQAVLGQFFFTPMKAATLMASMLRVDLGRGVRLDPGAGVSLTAALVD- 70  
Db 121 QALIG---PTIRGDKGQFFTP---KNLTDAIKILSPKGDKIIPACGTGGLSSCOAY 174  
Qy 71 -RLHTEPRDVAHVAVVAVETDPVVPYLRATLEECENAYGISYDLVEGDYL--LQOGAKLD 127  
Db 175 WELTYKOPTARYEILGDKDADMAWMLSSALLEISTNGFA---KVNDSLSLKFIDNPQYE 231  
Qy 128 GPFDLVIANPPYG-KLASDSLA-----RLATTARADVDPNVVYVAFWVRA 170  
Db 232 EQFDIVANPPFGTKIKVDNKAAILKDYQLGHSWKVENGTLCPSRHILGAQDPQILFIELC 291  
Qy 171 VISLKEQGRGVFIIPRSWANGPYRQFRHMLTAVSLDILHVPESRTKVPADTKVKQENV 230  
Db 292 VKLLKENGKRMALILPEG-----VFGKKSSEYVWEYLKNGRI 327  
Qy 231 IVAFSVRPOSSVVLRSVAHGEESTASSVPFSALVHDEDDDKI-----V 275  
Db 328 VPALIDCP-----RTTFQPYDTIKTNVLFKKTKEMPEKTQVAVAKRCGHDKRGRT 379  
Qy 276 HFAESASVPSAARFTLADLGIGVSTGKVVDFERNROYLDNLNDASGVWPMVYQSNIRSGKI 335  
Db 380 HYPGSLSPD---DFADIANLFXEG---IENRIWKSVCCLKKEYRVPYRYENDDDEAGKL 431  
Qy 336 DWPQVGARKQGFVAVEDVALRQLLPQGSYVVVKQTAKEDERRRRIAADVWGASRVAL-- 393  
Db 432 D--NIG-----QVITIGELIRMG---VLKIRKGHE-----VGSENYGTGDIFFIR 471  
Qy 394 DNKTYLHESQRPLEKNVARGLMNLNSTVLDQYFAPSFGHTQVNAVAGDL 442  
Db 472 TSDINNLFFSDP-----TNSVSEIYEMYSKKONIAAGDI 507

RESULT 6  
ABU23716  
ID ABU23716 standard; protein; 571 AA.  
XX AC ABU23716;  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #9243.  
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX OS Clostridium acetobutylicum.  
XX PN WO200277183-A2.  
XX PD 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US0009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX PI Wall D, Trawick JJ, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR N-PSDB; ACA27586.



CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX  
SQ Sequence 519 AA;  
Query Match 5.5%; Score 137; DB 6; Length 519;  
Best Local Similarity 23.1%; Pred. No. 0.00011;  
Matches 74; Conservative 57; Mismatches 124; Indels 66; Gaps 17;  
QY 27 GQFTPMKAAATLMASMLRVDDLRTGTVRLDPGAGVGSITAAALVDRLHTRPDVAHVAV 86  
Db 195 GEFPTQHSRLIA-QLAMHGQTSVKNLYDPAAGSGSL--LLOAKKHF-----AHII-- 244  
QY 87 ETDPEVVPYLRATLEECRN---AYGISYD---LVEGDYLLNQGAKLDGPPDLVIANPPY 139  
Db 245 -EDGFGQEIHTTTLARMNFMFLHNYDKFNILMGLNLTLPHTFGDDKPFDAIVSNPPY 303  
QY 140 GK--LASDLSA-----RLATTARAVDVNPNVYVAFVRAVISLKEQGRGVFI-VPRSWANG 191  
Db 304 SVKWIGSDPTLINDRFPAGVLAPKSKADPAFVLHLSYLSKSGRAAIVCFPGIFYRG 363  
QY 192 PYYRQFRHWM-----TAVSLDLHVFSRRTKVFAADTKVKQENVIYA 233  
Db 364 GAEQKIRKYLVDNNYVETVISLAPNLFPGTTIIVNLVLSKHT-----DTTQFIDASPL 419  
QY 234 FSVRQSSSVLRSVAHGEESIASVPFSAHVHDEDDDKIVHFAESASVPSAARTLAD 293  
Db 420 F--KKEINNNIITDN--HIEQIMQ-----VFDSKDD-VHEFAKSVSFEATA---AND 463  
QY 294 LGIGVSTGKVDVFRNRQYLTD 314  
Db 464 YNLSVSS--YVEARDNREVID 482  
RESULT 9  
ABP78365  
ID ABP78365 standard; protein; 510 AA.  
XX AC ABP78365;  
XX DT 07-MAR-2003 (first entry)  
XX DE N. gonorrhoeae amino acid sequence SEQ ID 3260.  
XX KW Antibacterial; infection; vaccine; gene therapy.  
XX OS Neisseria gonorrhoeae.  
XX PN WC200279243-A2.  
XX PD 10-OCT-2002.  
XX PF 12-FEB-2002; 2002WO-IB002069.  
XX PR 12-FEB-2001; 2001GB-00003424.  
XX PA (CHIR-) CHIRON SPA.  
XX  
SQ Sequence 514 AA;  
Query Match 5.5%; Score 137; DB 3; Length 514;  
Best Local Similarity 22.2%; Pred. No. 0.00011;  
Matches 73; Conservative 47; Mismatches 127; Indels 82; Gaps 14;  
QY 27 GQFTPMKAAATLMASMLRVDDLRTGTVRLDPGAGVGSITAAALVDRLHTRPDVAHVAV 86  
Db 196 GEFPTQHSRLIA-RLAVHGOEKVKNLYDPAAGSGSL-----LLOAKKQFDEHI--I 245  
QY 87 ETDPEVVPYLRATLEECRN---AYGISYD---LVEGDYLLNQGAKLDGPPDLVIANPPY 140  
Db 246 BEGFGQEIHTTTLARMNFMFLHNYDKFNILMGLNLTLPKLDKSKPFDAIVSNPPYS 305  
QY 141 -----KLASDLSARLATTARAVDVNPNVYVAFVRAVISLKEQGRGVFI-VPRSWAN 190  
Db 306 INWIGSDPTLINDR--RFAPAGVLAPKSKADPAFVLHLSYLSKSGRAAIVSFPPIFYR 363  
QY 191 GPYYRQFRHWM-----TAVSLDLHVFSRRTKV-FADTKVKQEN 229  
Db 364 GAEQKIRKYLVEGNYVETVIALPNLFYGTGIAVNLVLSKHKNDTIDQIFIDAS----- 418  
QY 230 VIVAFSVRQSSSVLRSVAHGEESIASVPFSAHVHDEDDDKIVHFAESASVP-----S 285  
Db 419 ---GFFKXETNNVLI-----EHHIAEIVKL-----PADKADVPHIAQN 454  
QY 286 AARFTLADLGIGVSTGKVDVFRNRQYLTD 314  
Db 455 AAQVQKONGYNLAVSSVVEAEDTREIID 483  
RESULT 8  
ABM70049  
ID ABM70049 standard; protein; 519 AA.  
XX AC ABM70049;  
XX DT 20-NOV-2003 (first entry)  
XX DE Photorhabdus luminescens protein sequence #3146.  
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
XX detection; food; gene expression; plant; animal; microorganism; toxin;  
XX antibiotic; biopesticide; virulence factor; disease model; plague;  
XX whooping cough.  
XX OS Photorhabdus luminescens.  
XX PN WC200294867-A2.  
XX PD 28-NOV-2002.  
XX PF 07-FEB-2002; 2002WO-IB003040.  
XX PR 07-FEB-2001; 2001FR-00001659.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
XX Buchrieser C;  
XX WPI; 2003-148459/14.  
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX Claim 2; SEQ ID NO 3146; 1205pp; French.  
XX The invention relates to the isolation of genes and their encoded  
XX proteins from Photorhabdus luminescens. The isolated sequences are  
XX sources of probes and primers for detecting the genome of P. luminescens  
XX and related species; to study polymorphisms; for gene analysis and for

[illegible]

XX AC ABU37303;  
XX DT 23-OCT-2003 (revised)  
XX DT 19-JUN-2003 (first entry)  
XX DE Protein encoded by Prokaryotic essential gene #22830.  
XX DE Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX KW Neisseria gonorrhoeae.  
XX OS WC200277183-A2.  
XX PN 03-OCT-2002.  
XX PF 21-MAR-2002; 2002WO-US009107.  
XX PR 21-MAR-2001; 2001US-00815242.  
XX PR 06-SEP-2001; 2001US-00948993.  
XX PR 25-OCT-2001; 2001US-0342923P.  
XX PR 08-FEB-2002; 2002US-00072851.  
XX PR 06-MAR-2002; 2002US-0362699P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zvekind JW;  
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX DR WPI; 2003-029926/02.  
XX DR N-PSDB; ACA41173.  
XX PT New antisense nucleic acids, useful for identifying proteins or screening  
XX PT for homologous nucleic acids required for cellular proliferation to  
XX PT isolate candidate molecules for rational drug discovery programs.  
XX PS Claim 25; SEQ ID NO 65227; 1766pp; English.  
XX CC The invention relates to an isolated nucleic acid comprising any one of  
XX CC the 6213 antisense sequences given in the specification where expression  
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid  
XX CC encoding a polypeptide whose expression is inhibited by the antisense  
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX CC polypeptide or its fragment whose expression is inhibited by the  
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding  
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX CC proliferation or the activity of a gene in an operon required for  
XX CC proliferation; (7) identifying a compound that influences the activity of  
XX CC the gene product or that has an activity against a biological pathway  
XX CC required for proliferation, or that inhibits cellular proliferation; (8)  
XX CC identifying a gene required for cellular proliferation or the biological  
XX CC pathway in which a proliferation-required gene or its gene product lies  
XX CC or a gene on which the test compound that inhibits proliferation of an  
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX CC compound's activity; (11) a culture comprising strains in which the gene  
XX CC product is overexpressed or underexpressed; (12) determining the extent  
XX CC to which each of the strains is present in a culture or collection of  
XX CC strains; or (13) identifying the target of a compound that inhibits the  
XX CC proliferation of an organism. The antisense nucleic acids are useful for  
XX CC identifying proteins or screening for homologous nucleic acids required  
XX CC for cellular proliferation to isolate candidate molecules for rational  
XX CC drug discovery programs, or for screening homologous nucleic acids  
XX CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX CC the target prokaryotic essential genes. Note: The sequence data for this  
XX CC patent did not form part of the printed specification, but was obtained  
XX CC in electronic format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences. (Updated on 23-OCT-2003 to  
XX CC standardise OS field)  
XX CC Sequence 514 AA;

Query Match 5.5%; Score 135; DB 6; Length 514;  
Best Local Similarity 22.0%; Pred. No. 0.00016;  
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;  
QY 27 GQFETPMKAATLMASMLRVDDLRGTVRVLPDGAGVGSILTAALVDRLHTRPDAVHVAV 86  
DB 196 GEFPTQSVSKLIA-RLAVHGOEKVKNKIYDPACGSSL-----LLQAKKQFDEHI--I 245  
QY 87 ETDFPVVYLRATLECRN---AYGISYD---LVGYLLNQAKLOGPFDLVIANPPYG 140  
DB 246 EEGFFGOEINHNTTYNLRMMFLHVNYNKPHIELGDTLTNPCLKSKPFDVAVVSNPPYS 305  
QY 141 -----KLASDLARLATTARADVDPVYVAVFWRVAVISLKEQSGVFI-VPRSWAN 190  
DB 306 IDWIGSDDDPTLINDD--RFAPAGVLAPKSKADPAFILLHALNLYSGRGAIAVSPFGIYR 363  
QY 191 GPYYRQFRHLM-----TAVSLDILHVFESRTKVPADTKVQENVIV 232  
DB 364 GGAQKIRQYLVVEGNYVETVIALAPNLFYGTCTIAVNILVLSKHK-----DNTDI 412  
QY 233 AF-----SVRPQSSSVLSRSVAHGESIASVSPFSALVHDEDDDKIVHFAESAVP--- 284  
DB 413 QFIDASGFFKKTNNVLT-----EEHIAEIVKL-----FADKADVPHIA 452  
QY 285 -SAARFTLADLIGVSTGKVVDFFNRQYLTLD-----NLDASGVVPMV 325  
DB 453 QNAAQOTVKDNGYNLAVSSVVEAEDTREVIDRQLNAEISETVAKI 498  
RESULT 12  
ID AAB07560 standard; protein; 2162 AA.  
XX AC AAB07560;  
XX DT 20-OCT-2000 (first entry)  
XX DE Protein encoded by the bleomycin (BLM) gene cluster ORF30.  
XX KW BLM gene cluster; bleomycin gene cluster; polyketide metabolite;  
XX KW bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;  
XX KW thiazoline; bithiazoline; microbial metabolite; sugar.  
XX OS Streptomyces verticillus.  
XX PN WO200040704-A1.  
XX DT 13-JUL-2000.  
XX PF 06-JAN-2000; 2000WO-US0000445.  
XX PR 06-JAN-1999; 99US-0115435P.  
XX PR 05-FEB-1999; 99US-0118848P.  
XX PR 05-JAN-2000; 2000US-00477962.  
XX PA (REGC) UNIV CALIFORNIA.  
XX PI Shen B, Du L, Sanchez C, Chen M, Edwards DJ;  
XX PI WPI; 2000-465974/40.  
XX PN N-PSDB; AAA58471.  
XX CC New bleomycin gene cluster components useful for peptide and/or  
XX CC polyketide metabolites, especially bleomycin, production and for  
XX CC chemically modifying biological molecules.  
XX CC Disclosure; Page 97-137; 162pp; English.  
XX CC AAB07566-78 represent proteins encoded by open reading frames (ORFs) 8 to  
XX CC 30 of the BLM (Bleomycin) gene cluster. The proteins encoded by the gene  
XX CC cluster are useful for producing peptides and/or polyketide metabolites,  
XX CC especially bleomycin or bleomycin analogues. They are also useful for  
XX CC chemically modifying biological molecules to produce branched methyl





Search completed: October 1, 2004, 16:12:55  
Job time : 67.6786 secs

P Lank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:09:01 ; Search time 19.8245 Seconds  
(without alignments)  
1265.619 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHP1ASTETRRQAALGKLDP.....QETLDVAVARLFCIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PGTUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Match	Length	DB ID	Description
1	136.5	5.5	1309	4	US-09-975-413A-10	Sequence 10, Appl
2	122	4.9	534	4	US-09-107-532A-6592	Sequence 6592, Ap
3	120.5	4.9	535	4	US-09-107-532A-6593	Sequence 6593, Ap
4	117.5	4.8	507	4	US-09-252-991A-26552	Sequence 26552, A
5	116.5	4.7	414	4	US-09-252-991A-21182	Sequence 21182, A
6	114.5	4.6	597	4	US-09-252-991A-31010	Sequence 31010, A
7	113.5	4.6	1009	4	US-09-693-146-4	Sequence 4, Appl
8	111.5	4.5	549	4	US-09-693-146-2	Sequence 2, Appl
9	109.5	4.4	672	4	US-09-252-991A-32554	Sequence 32554, A
10	109	4.4	1007	4	US-09-957-005-9	Sequence 9, Appl
11	106.5	4.3	504	4	US-09-134-001C-2980	Sequence 2980, Ap
12	106	4.3	406	4	US-09-071-035-114	Sequence 114, App
13	105	4.3	557	4	US-09-134-000C-4354	Sequence 4354, Ap
14	104.5	4.2	515	4	US-09-489-039A-10635	Sequence 10635, A
15	101	4.1	202	4	US-08-311-731A-13	Sequence 13, Appl
16	100.5	4.1	296	4	US-09-134-001C-3320	Sequence 3320, Ap
17	99.5	4.0	338	4	US-09-489-039A-11940	Sequence 11940, A
18	99.5	4.0	604	2	US-08-746-283-32	Sequence 32, Appl
19	99.5	4.0	604	2	US-08-746-257A-30	Sequence 30, Appl
20	99.5	4.0	605	1	US-08-333-802-2	Sequence 2, Appl
21	99.5	4.0	605	4	US-09-347-878-46	Sequence 46, Appl
22	99	4.0	528	4	US-09-071-035-44	Sequence 44, Appl
23	99	4.0	547	4	US-09-071-035-42	Sequence 42, Appl
24	98.5	4.0	513	4	US-09-489-039A-14224	Sequence 14224, A
25	98.5	4.0	554	4	US-09-540-236-3778	Sequence 3778, Ap
26	98.5	4.0	618	3	US-09-299-378-4	Sequence 4, Appl
27	97.5	3.9	301	4	US-09-489-039A-10251	Sequence 10251, A

28 96.5 3.9 926 4 US-09-252-991A-32551 Sequence 32551, A

29 95 3.8 800 4 US-09-328-352-5709 Sequence 5709, Ap

30 95 3.8 3241 4 US-03-841-786-1 Sequence 1, Appl

31 95 3.8 3892 4 US-09-328-352-5503 Sequence 5503, Ap

32 94.5 3.8 287 4 US-09-328-352-7035 Sequence 7035, Ap

33 94.5 3.8 336 3 US-09-105-537-27 Sequence 27, Appl

34 94.5 3.8 351 4 US-09-134-000C-4436 Sequence 4436, Ap

35 94.5 3.8 11877 3 US-09-105-537-6 Sequence 6, Appl

36 93.5 3.8 398 4 US-09-134-001C-5523 Sequence 5523, Ap

37 93.5 3.8 927 4 US-09-841-786-3 Sequence 3, Appl

38 93.5 3.8 3074 4 US-09-543-681A-5508 Sequence 5508, Ap

39 93 3.8 253 4 US-09-180-109A-35 Sequence 35, Appl

40 93 3.8 254 4 US-09-180-109A-31 Sequence 31, Appl

41 93 3.8 664 4 US-09-328-352-6193 Sequence 6193, Ap

42 93 3.8 828 4 US-09-198-452A-323 Sequence 323, App

43 92.5 3.7 197 4 US-09-252-991A-27889 Sequence 27889, A

44 92.5 3.7 414 4 US-09-252-991A-27975 Sequence 27975, A

45 92.5 3.7 419 3 US-08-258-287B-57 Sequence 57, Appl

## ALIGNMENTS

## RESULT 1

US-09-975-413A-10

; Sequence 10, Application US/09975413A

; Patent No. 6593122

; GENERAL INFORMATION:

; APPLICANT: Maunus, Robert

; APPLICANT: Xu, Shuang-Yong

; APPLICANT: Benner, Jack

; TITLE OF INVENTION: Method For Cloning And Expression of BserI Restriction Endonuclease

; TITLE OF INVENTION: BserI Methylase In E. coli

; FILE REFERENCE: NEB-193

; CURRENT APPLICATION NUMBER: US/09/975,413A

; CURRENT FILING DATE: 2001-10-11

; NUMBER OF SEQ ID NOS: 45

; SOFTWARE: Patent version 3.1

; SEQ ID NO 10

; LENGTH: 1309

; TYPE: PRT

; ORGANISM: Bacillus species R (CAMB2669)

US-09-975-413A-10

Query Match 5.5%; Score 136.5; DB 4; Length 1309;

Best Local Similarity 24.4%; Pred. No. 9.9e-05;

Matches 63; Conservative 40; Mismatches 94; Indels 61; Gaps 13;

QY 17 KLDPTTQAVLIGQFPTPKAATLWASMLRVDLRLGTV--RVLDPGAGVGSILTAALVDRLHT 74

Db 630 QINPERQKRLQVFTSGPLAELIATF-----AEGSTASSVIDPMCGGOMLTA-VNSINS 683

QY 75 ERPDVAHVHVAVETDPVPVYLRATLEECRNAYG----ISVDLVEGD-YLLNQAKLD-GP 129

Db 684 K-----ANLSGIDIDPI-----AMNKCIDRLGNKKSLDLIGSAFSWNTIKQLKLS 731

QY 130 FDLVIANPPI-----GKLSDSLARLATTAR-----AVD-----158

Db 732 FDLVITNPVRYQSLSSKLEGLVDLPDSETVRNDLLEVVSQLDHLHRDKVEFTVKS 791

QY 159 ---VPNVYAFVWRAVISLKEQGRGVFIIVPRSWANGPPYRQFHEMLTAVSLDLHVFS 215

Db 792 YGSLSDLAVPSILCAMLTSVGHLAMVVPESWLNRYAHPHYLLLLFKIKWVVDYN 851

QY 216 RTKVFADTKVKQENVIVA 233

Db 852 RT-WFKDAQVK-TNLVVA 867

## RESULT 2

US-09-107-532A-6592

; Sequence 6592, Application US/09107532A

; Patent No. 6563275

GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Deneka  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 6592:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...534  
SEQUENCE DESCRIPTION: SEQ ID NO: 6592:  
US-09-107-532A-6592

Query Match 4.9%; Score 122; DB 4; Length 534;  
Best Local Similarity 24.1%; Pred. No. 0.00075;  
Matches 70; Conservative 35; Mismatches 116; Indels 70; Gaps 15;  
QY 27 GQFFPMKAATMASMLRVDDL-RGTVRLDPGAGVGSLSAALVDLHTRPD-VAVHV 84  
Db 198 GEFYTHPMVDMMAQIVTLDKERRFFSVFDTMTSGSL--MLNVRNYLTHPDNKKYHQ 255  
QY 85 AVETPFVVPYLRATLEECRNAYGI---SYDLVEGDYLLNOGAKLDGP--FDLVIANPP- 138  
Db 256 ELNNTTYNLAKNMLIL-----HGVDAEEMNLRNGD-TLNKQWPTDEPTTFDAVVMNPPY 308  
QY 139 -----YGLKASDSLARLATTARAVDVPNVYVAFWVRVAVISLKEQGRG 180  
Db 309 SANWSADITFLDSDRENRYGKLAPKSKADF-----AFLHGFVHLKETGTM 354  
QY 181 VFIWPSWANGPYRQFRHMLMTAVSLD---ILHVFESRTKVPADTKVQENVIVAFSVR 237  
Db 355 AIVLP-----HGVLFGRGAEGVIRQKLEDGSIYAVIGMPANLFFGTISI--PTTVIVLKN 408  
QY 238 PQSSSVVL---SRVAHG-----PESTASSVPFSAVHDEDDDKIVHFA 278  
Db 409 RQTRDVLFDASREFVKGKQNKLSSENIQKIL--ETVAKDKVKAHLA 457

RESULT 3

US-09-107-532A-6593  
; Sequence 6593, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneka  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 6593:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 535 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Enterococcus faecium  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...535  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6593:  
US-09-107-532A-6593

Query Match 4.9%; Score 120.5; DB 4; Length 535;  
Best Local Similarity 23.0%; Pred. No. 0.0011;  
Matches 79; Conservative 39; Mismatches 135; Indels 91; Gaps 18;  
QY 27 GQFFPMKAATMASMLRVDDL-RGTVRLDPGAGVGSLSAALVDLHTRPD-VAVHV 84  
Db 198 GEFYTHPMVDMMAQIVTLDKERRFFSVFDTMTSGSL--MLNVRNYLTHPDNKKYHQ 255  
QY 85 AVETPFVVPYLRATLEECRNAYGI---SYDLVEGDYLLNOGAKLDGP--FDLVIANPP- 138  
Db 256 ELNNTTYNLAKNMLIL-----HGVDAEEMNLRNGD-TLNKQWPTDEPTTFDAVVMNPPY 308  
QY 139 -----YGLKASDSLARLATTARAVDVPNVYVAFWVRVAVISLKEQGRG 180  
Db 309 SANWSADITFLDSDRENRYGKLAPKSKADF-----AFLHGFVHLKETGTM 354  
QY 181 VFIWPSWANGPYRQFRHMLMTAVSLD---QFRHMLMTAVSLDILHVFESRTKVPADTKVQENVIVAFS 235  
Db 355 AIVLP-----HGVLFGRGAEGVIRQKLEDGSIYAVIGMPANLFFGTISI--PTTVIVL 406  
QY 236 VPQSSSVVL---SRVAHGEESIASVPFSAVHDEDDDKIVH-FAESASVPSAARFTL 291  
Db 407 KNRQTRDVLFDASREFVKGKQNKLS-----EENIQKILETYAKDKVKAHLAT 458

QY 292 ADLIGVSTGKVDPRNQY-----LTDNLDASGVPMVYQSN 329  
Db 459 FD-----BIKENDYLNIPRYVDTFEEBPDIMVHGN 491

## RESULT 4

US-09-252-991A-26552  
; Sequence 26552, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 26552

; LENGTH: 507  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-26552

Query Match 4.8%; Score 117.5; DB 4; Length 507;  
Best Local Similarity 24.3%; Pred. No. 0.002;

Matches 107; Conservative 52; Mismatches 196; Indels 95; Gaps 23;

QY 50 GTVRVLDPCAGVGLTAALVDRLHTRPDVAVHVAVETDPFVVPVLRATLEECRNAYGI 109

Db 64 GRMKICDPCAGEG---VALAAEAHILGRD-KVQALAVEYDRADHARGLLERVLHS--- 116

QY 110 SYDLVEGYLLNOGAKLDGDFDLVIANPPYKGLASDLARLATTARAVDPNYYVAFWVR 169

Db 117 --DLF--DTMISRQS-----FGLLWLNPPYGLDLVADHSG--ASQYQSGRRRLKAFYQR 165

QY 170 AVISLKEQGGGVFVPRSWANGPY--RQFRHWTAVSLDLHVFESRTKVFAATKVK 226

Db 166 CULPLQYGMVLLIV-----PHYVLDDELTCWLSN-----HFTGLRMVAAADPTFK 211

QY 227 QENVIVAFSVRPO-----SSSWLSRSVAHG-----EESIASVSPFSALVHDEDDDKI 274

Db 212 Q---VVIFGIRVRQDLARADANQVSRLOAQAGQKAEETPAAPWPEYVVLPAATSEL 268

QY 275 VIFASASVPSA-----ARFTLADLIGVSTGKVDPRNQYLTNDLNDASGVV 322

Db 269 EHFYRVTLPEQFAGEIORLGLWDPFNLHFAQAGLQPRPPVRELSRWHLALAL-AAGAI 327

QY 323 PMYQSNIRSGKI-----DWPQVGARKPGGFVAVEDVALRQLLPQGSYVVVKRQTAKER 377

Db 328 SGVRS--KSGRILVKGDTYKDKARKTE-FTEDED-----GNITEVRILT---DR 372

QY 378 RRVIAAVND-----GASRV-ALDNKNTYLSQSOREPLEKNVARGMLNLSVTLDQYFPA 430

Db 373 FIPIIRAWMTSPSSVQNGRVLTISSAAATEAEPEQPEAPEPELDFSPGQVVM-----TA 428

QY 431 FSGHTQVWAGDRLRULPFLCR 450

Db 429 AVSHL-VETGQLNPAPLNR 447

## RESULT 5

US-09-252-991A-21182

; Sequence 21182, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21182  
; LENGTH: 414

; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21182

Query Match 4.7%; Score 116.5; DB 4; Length 414;  
Best Local Similarity 24.4%; Pred. No. 0.0019;

Matches 77; Conservative 31; Mismatches 106; Indels 101; Gaps 16;

QY 17 KLDPTTQAVLGQF--FTPDKAATLMASM-----LRVDD 47

Db 22 RLDPPSEARLGVLPSPFIPANQETSMALMFPRLARNFAKNGYYPTDEPTLERALNALMPSD 81

QY 48 LRGTVRVLDPCAGVGLTAALVDRLHTRPDVAVHVAVETDPFVVPVLRATLEECRNAY 107

Db 82 --GPMCLDPCAGEGVALAEASHALGSEQ---AKAFAVEFDAERARHARGLVHCUHA- 134

QY 108 GISYDLVEGYLLNOGAKLDGDFDLVIANPPYKGLASD-----SLARLATTARAV 157

Db 135 ---DLM--DTMISKQS-----FGLLWLNPPYGLSKDVNGNIGYQCGGARLEK----- 178

QY 158 DVENVVAFWVRVAVISLKEQGGGVFVPR-----SWANGPYVYQFRHWTAVSLDI 209

Db 179 -----LFQRSLSLQYSGVLVFIQVGVLDLAEVLGMLT-----RHY-----TDLRI 220

QY 210 LHVFESERTK--VFADTKVKQENVIVAFSVRPSQSSVV--LSRSVAHGE---ESIASVVPF 262

Db 221 YRAVETQFKQVIFGRRVRQR-----ELAPGCVKAVRNLLQVGLGEVAAEELPSEWPF 274

QY 263 SALVHDEDDDKIVHF 277

Db 275 LPYIVPASPAEPEHF 289

## RESULT 6

US-09-252-991A-31010

; Sequence 31010, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31010

; LENGTH: 597  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31010

Query Match 4.6%; Score 114.5; DB 4; Length 597;  
Best Local Similarity 23.4%; Pred. No. 0.0056;

Matches 44; Conservative 30; Mismatches 75; Indels 39; Gaps 9;

QY 27 GQFTPMKAATLMASMLRVD---DLRGTVRVLDPCAGVGLTAALVDRLHTE-----R 76

Db 175 GEFYTPTRAITAFMAD--RVNPRLDKRET--VMDPACGTGGFLTAIDHFRNQLSKSSAE 230



QY 198 -----RHLMATAVS-----LDI 209  
Db 217 LGELTHYIGIEDITONKLAQOHIEKLAAYRDNLGLANRHYFIGALEBLESSGDRPLSL 276  
QY 210 LHVFESEKRVKPADTKVQ--ENVIVAFVRPOSS-----SVVLSRSVAHG 252  
Db 277 LLVDIDNFRINDSLGHOTGDKLLVSLARRLRSLCLGDGATLARFASNEFAVLLDDTAVEK 336  
QY 253 EESIASVPFSAVHDEDDK--DDKIVHFAESASVPSAARF-----TLADLGIGVSTG 301  
Db 337 GESIAAQV-----LHMLDKPLFVDNQLNITGSLASAPQHGCDPQTLMKYAGLALHKA 391  
QY 302 KVDPRNROYLTDNLASGVPMVYQSN-----IRSGK-----I 335  
Db 392 KANGKHQVQVTEALTAEASYKLFVESNLRRALAQNELAVHYQPKCLRSQQLGLEALL 451  
QY 336 DW--POVGARPKQGVAV-EDVALRQLLPQGSYVVVVKQTAKEDRRRRVIAAVWGASRVA 392  
Db 452 RWQHPKGMIRPDRFISVAETGL--IVPIGKWI--REACRQARELAEGAL--GELQIA 505  
QY 393 LD-----NKTNYLHESORP--LEKNVARGMLMLNSTVLDOYFRAFS-----432  
Db 506 INLSPKQFTDPDLVGSIAALHEENIPASQLELELSTESLLDATTDDTRQQLERKLSGLT 565  
QY 433 -----GHTQVWAGDLRLPFLCREDLILL-----AKVVPDGLPQDETLDVAV-----475  
Db 566 LAMDDFGTGYSSLSY--LKKFPI--DVIKIDRSFIKIDPDSODDMEITSAVIAMAHNL 619  
QY 476 --RLFCPEIPESA 485  
Db 620 KLVVVAEGVESA 631

## RESULT 10

US-09-957-005-9  
; Sequence 9, Application US/09957005  
; Patent No. 6596524  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Zhenyu  
; APPLICANT: Zhou, Jing  
; APPLICANT: Xu, Shuang-yong  
; TITLE OF INVENTION: Method For Cloning And Expression of BsmAI Restriction  
; FILE REFERENCE: NEB-192  
; CURRENT APPLICATION NUMBER: US/09/957,005  
; PRIOR FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 20  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 9  
; LENGTH: 1007  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus A664  
US-09-957-005-9

Query Match 4.4%; Score 109; DB 4; Length 1007;  
Best Local Similarity 18.6%; Pred. No. 0.051;  
Matches 105; Conservative 102; Mismatches 215; Indels 142; Gaps 28;  
QY 17 KLDPTTQAVLGQFFTPMKAATLM-----ASMLRVDDLRTGTVRLDPGAGVGSLSAALVD- 70  
Db 36 QMDPLERKETSFTALELTDVMMQBLVSYLLKSDKIDTELKFLPCVGTGNGVNFVSYLKE 95  
QY 71 ----RLHTERPDAVHVAWE-----TOPFVVPVYLRATL-EECRNAYGISYD 112  
Db 96 ISKLQLHKEQIETLNNIYVADINQALLEYKLLSKFAKLYPDIDLSEYFNHSGSAL 155  
QY 113 LVE-----GDLNQAQKLDGPFDLVIANPPY-----GKLASD-----145  
Db 156 LIDVAEQEYIKITDVFDEVVEK-----FDIVVTPPPYKLNKAEKGQVSNLDVEYI 209  
QY 146 SLARLATTARV-----DVPNVYVAFWVRVAVISLKEQGRGV--FIVPRSWANGPPY 194

Db 210 DRARYAEIKKVKRIFNYSTDGLNLKLF-VEEIIDKVANPNGFVSLLIPSSILTDKTC 268  
QY 195 RQFR-HWMTAVSLDILHVFESRTKVFPADTKVQENVIKQENIVAFSVRPOSSSVLSRSVAHGE 253  
Db 269 TKLRTHMLVDSNILLSIKMINEGSGVIDA---QQALSAILIQKGRKTESIKVTKDYNNP 324  
QY 254 ESIASSVPFSAVHDEDDK--VHFABASVPSAARF-TLADLGIGVSTGVKVDPRNRQ 310  
Db 325 NQI--TDINMEDILLNENTGNALFAINNHEYFLKQLRKFPPVVKOLDP-----IINLRGEL 377  
QY 311 YLTNLDASGVPMVYQSNIRSGKIDMPQVGARKPQGVAVEDVALROLL-----PQ 362  
Db 378 DLTANKS--IV-----NIDTG--YPLLRGN-IGYIELDTCSGFEVSKDFIENSKK 425  
QY 363 GSYVVVVR-----QTKEDRRRVIAAVWGASRVALDNKTNLYHESQRPLEKKNVARGLM 416  
Db 426 SRYIKEKRIVCQVVNNMKERRVTFALV--EENVVLGNSCNFI--SVMDNDYNDILYAI 480  
QY 417 LWL-NSTVLDOYFRAFSGHTQVWAGDLRLP-----FLCREDLILIAK 458  
Db 481 LGLNTSIIINWLFKLTSSNNHNNVEIDCFVPPIGSPYLNKISNLVKKYLKNSKOSLLEK 540  
QY 459 V-----VPDGLPQDETLDV 473  
Db 541 IEEVAYIAYGIREAKEDNEKDDI 564

## RESULT 11

US-09-134-001C-2980  
; Sequence 2980, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 2980  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2980

Query Match 4.3%; Score 106.5; DB 4; Length 504;  
Best Local Similarity 27.1%; Pred. No. 0.029;  
Matches 57; Conservative 30; Mismatches 80; Indels 43; Gaps 16;  
QY 4 IASTETRRQAAALG-----KLDPTTQAVLGQFFTPMKAATLMASMLRVDDLRTGTVRL 55  
Db 137 VGTESRKQDVLRGVVEYFIAPASAGKNAGFEYTPSSIVKLLVEM--IEPYKG--RIY 192  
QY 56 DFCAGVGSLS---TAALVDRLHTERPDAVHVAWEVDPFVVPVYLRATLECRNAYGISYD 112  
Db 193 DPCGSGGMFVQSERFVEKHQGLDDIAY--GOESNPTTWKLAQNLA-----IRGIDND 246  
QY 113 LVE--GDYL---LNQAKLDGPFDLVIANPPYKGLASD-SLARLATTAR-AVDVP---NV 162  
Db 247 LGERNSDTFPHDLHKLGA---DYILANPPFN--ASDWGQERLLDDYRWQFGIPEKGA 300  
QY 163 YVAFWVRVAVIS-LKEQGRGVFIVPRSWANG 191  
Db 301 NYA-WIEHMISKLAENGTAAGFVL---ANG 325

## RESULT 12

US-09-071-035-114  
; Sequence 114, Application US/09071035

Patent No. 6448043  
GENERAL INFORMATION:  
APPLICANT: Gil H. Choi  
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
NUMBER OF SEQUENCES: 496  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/071,035  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB369P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 114:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 406 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-071-035-114

Query Match 4.3%; Score 106; DB 4; Length 406;  
Best Local Similarity 23.0%; Pred. No. 0.023;  
Matches 102; Conservative 66; Mismatches 170; Indels 106; Gaps 25;  
QY 33 MKAATLMASMLRVDDLRGTGVRLDPGAGVGSILTAALVDRLHTRPDVAVHV-----AV 86  
Db 1 MKKLKMMGIMLFV-----STVLV-----GCGTTAETKIDKATEKTSVSKKVLNLMENSEI 51  
QY 87 ETDPFVVPVYLRTALEECRNAYGISYDLVEGDYLLNQAK-----LDGPFDLVIANPPYG 140  
Db 52 GMSDSIFTQDEASINAQSNVFEGLYQLDEKQDLIPAAAKEMPEISEDGK-RYTIKLREDG 110  
QY 141 KLASDLSLARLATTARAVDPNVVYAFWVRVAVISLKEQGRGVFIVPRSWANGPYRQFRHW 200  
Db 111 KWSNGD-----AVTA-----NDVFPAW-RKLANPKNQANYFFLEGTILNG-----150  
QY 201 LMTAVS-----LDLHVFESRTKVFADTKVKQENVIVAFS-VRPQSSSVLSRSVAH 251  
Db 151 --TAITKEKAPBELGVKALDDYTLVLETKPVYFTSLIAFPFPQNEAFVKEKGQAY 208  
QY 252 G--EESTASSVPF-----SALVHD-----EDDDKI-----VHF-----AESASVPSAAR 288  
Db 209 GTSEMIVSNGPFLMKWQDSAMSDFVRNPYYDKVKSETIHFVLKETTIVNLT--266  
QY 289 FTLADLIGIGYSTGKVDV--RNRQYLTNDLDASGV--PMVYQSNIRSGKIDWPQVGA---342  
Db 267 YESGELDVAVLTG---DPAKQNR-----DNPDEAIEERSKVYSRLNQKRNKPSIFANEN 319  
QY 343 -RKPQGFVAVEDVALRQLLPQGS-----YV-----VVKROTAKEDRRRVAIAVWDGASRVA 392  
Db 320 VRKALAYALDKKSLVDNILDGSKSKEIYGIPEKVFVNPETNEDFRQAGALVKTDAKKA-378  
QY 393 LDNKNLYLHESQRPLEKNVARGLM 416

Db 379 ----KEYLDKAKAELNGDVAIELL 398  
RESULT 13  
US-09-134-000C-4354  
Sequence 4354, Application US/09134000C  
Patent No. 6617156  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 032796-032  
CURRENT APPLICATION NUMBER: US/09/134,000C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/055,778  
PRIOR FILING DATE: 1997-08-15  
NUMBER OF SEQ ID NOS: 6812  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 4354  
LENGTH: 557  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
US-09-134-000C-4354

Query Match 4.3%; Score 105; DB 4; Length 557;  
Best Local Similarity 23.0%; Pred. No. 0.05;  
Matches 102; Conservative 66; Mismatches 170; Indels 106; Gaps 25;  
QY 33 MKAATLMASMLRVDDLRGTGVRLDPGAGVGSILTAALVDRLHTRPDVAVHV-----AV 86  
Db 8 MKKLKMMGIMLFV-----STVLV-----GCGTTAETKIDKATEKTSVSKKVLNLMENSEI 58  
QY 87 ETDPFVVPVYLRTALEECRNAYGISYDLVEGDYLLNQAK-----LDGPFDLVIANPPYG 140  
Db 59 GMSDSIFTQDEASINAQSNVFEGLYQLDEKQDLIPAAAKEMPEISEDGK-RYTIKLREDG 117  
QY 141 KLASDLSLARLATTARAVDPNVVYAFWVRVAVISLKEQGRGVFIVPRSWANGPYRQFRHW 200  
Db 118 KWSNGD-----AVTA-----NDVFPAW-RKLANPKNQANYFFLEGTILNG-----157  
QY 201 LMTAVS-----LDLHVFESRTKVFADTKVKQENVIVAFS-VRPQSSSVLSRSVAH 251  
Db 158 --TAITKEKAPBELGVKALDDYTLVLETKPVYFTSLIAFPFPQNEAFVKEKGQAY 215  
QY 252 G--EESTASSVPF-----SALVHD-----EDDDKI-----VHF-----AESASVPSAAR 288  
Db 216 GTSEMIVSNGPFLMKWQDSAMSDFVRNPYYDKVKSETIHFVLKETTIVNLT--273  
QY 289 FTLADLIGIGYSTGKVDV--RNRQYLTNDLDASGV--PMVYQSNIRSGKIDWPQVGA---342  
Db 274 YESGELDVAVLTG---DPAKQNR-----DNPDEAIEERSKVYSRLNQKRNKPSIFANEN 326  
QY 343 -RKPQGFVAVEDVALRQLLPQGS-----YV-----VVKROTAKEDRRRVAIAVWDGASRVA 392  
Db 327 VRKALAYALDKKSLVDNILDGSKSKEIYGIPEKVFVNPETNEDFRQAGALVKTDAKKA-385  
QY 393 LDNKNLYLHESQRPLEKNVARGLM 416  
Db 386 ----KEYLDKAKAELNGDVAIELL 405

RESULT 14  
US-09-489-039A-10635  
Sequence 10635, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27

```

; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 10635
; LENGTH: 515
; TYPE: PR1
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-10635

```

Query Match	4.2%;	Score 104.5;	DB 4;	Length 515;
Best Local Similarity	19.1%;	Pred. No. 0.05;		
Matches	53;	Conservative 38;	Mismatches 73;	Indels 113; Gaps 12;
QY	16	GKLDPT---TQAVLGOFFFTPMKAATLMASMLR--VDDLRTGTVRLDPGAGVGSLSAALVD	70	
Db	133	GLIDKSAQDTRSGAGQYFTPSRLVNTIVRLIQPSLSEL-----IQDPAIGSGGF-----	181	
QY	71	RLHTRPDVAHVHVAYETDPFVVPYLRLATLEECNAYG-----	108	
Db	182	-----LVSADSGFI-----RNKYSHKYEOTNPKYQGVIEKNTRR	216	
QY	109	-----ISYDLVEG DYLLNQGAKLDGPFOLVIANPPYGYKLASDSLRLATTAAVD	158	
Db	217	ICLMNTFLRLDANIYGALTDDAELD-KADVIVANPPFGNKGQRPURS-----D	269	
QY	159	VP-----NVTVAFWRAVISLKEQGRGVFIIVPRSWANGPYRQFRHMLMTAVSLDILH---	211	
Db	270	IPFPNTNKQLAFLQHIYGLRAGGRAAVLPDNL-----FESGVTGTVDRDLNNKCN	322	
QY	212	-----VFESR-----TKVFADTKVKQEN	229	
Db	323	LHTLRLPTGIFYAHGVKTNVLFFFTGTGAKNKYQOEN	359	

RESULT 15

US-08-311-731A-13

Sequence 13, Application US/08311731A

Patent No. 6583266

GENERAL INFORMATION:

APPLICANT: SMITH, DOUGLAS

APPLICANT: MAO, JEN-I

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR

TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 411

CORRESPONDENCE ADDRESS:

ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.

STREET: 600 ATLANTIC AVENUE

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: USA

ZIP: 02210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/311,731A

FILING DATE:

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: GATES, EDWARD R.

REGISTRATION NUMBER: 31,616

REFERENCE/DOCKET NUMBER: C0044/7125

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/720-3500

TELEFAX: 617/720-2441

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 202 amino acids

TYPE: amino acid

TOPOLOGY: linear

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: October 1, 2004, 16:13:07 ; Search time 103.928 Seconds  
(without alignments)  
1504.830 Million cell updates/sec

Title: US-10-668-047-2

Perfect score: 2469

Sequence: 1 VHPASTETTRQAALGKLDP.....QETLDVAVARLFCPEIPESAS 486

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubppaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubppaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubppaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223	9.0	1068	10	US-09-966-997-11
2	164	6.6	543	12	US-10-335-977-7269
3	161	6.5	543	12	US-10-335-977-7268
4	141	5.7	571	12	US-10-282-122A-51640
5	135	5.5	514	12	US-10-282-122A-55227
6	134	5.4	514	12	US-10-282-122A-65854
7	134	5.4	534	12	US-10-282-122A-72313
8	133.5	5.4	381	12	US-10-335-977-7481
9	133.5	5.4	384	12	US-10-335-977-7482
10	133.0	5.3	545	10	US-09-882-227-600
11	126	5.1	545	12	US-10-335-977-7271
12	125.5	5.1	317	9	US-09-712-363-185
13	125.5	5.1	317	12	US-10-282-122A-62340
14	125.5	5.1	317	12	US-10-282-122A-64484
15	124.5	5.0	295	12	US-10-282-122A-53706

16 124.5 5.0 511 9 US-09-815-242-10737 Sequence 10737, A

17 124.5 5.0 530 12 US-10-282-122A-56854 Sequence 56854, A

18 122.5 5.0 491 12 US-10-335-977-73395 Sequence 7395, Ap

19 122.5 5.0 1253 12 US-10-335-977-7339 Sequence 7399, Ap

20 122 4.9 358 10 US-09-882-227-164 Sequence 164, App

21 122 4.9 530 12 US-10-282-122A-57917 Sequence 57917, A

22 121.5 4.9 537 9 US-09-486-734A-8 Sequence 8, Appli

23 121.5 4.9 591 12 US-10-335-977-7396 Sequence 7396, Ap

24 120 4.9 348 12 US-10-282-122A-76713 Sequence 76713, A

25 119 4.8 823 9 US-09-895-913A-244 Sequence 244, App

26 118 4.8 291 12 US-10-282-122A-72248 Sequence 72248, A

27 117 4.7 518 12 US-10-282-122A-44252 Sequence 44252, A

28 117 4.7 550 12 US-10-335-977-7272 Sequence 7272, Ap

29 116 4.7 317 9 US-09-815-242-13654 Sequence 13654, A

30 116 4.7 317 10 US-09-769-744A-116 Sequence 116, App

31 116 4.7 317 12 US-10-282-122A-74193 Sequence 74193, A

32 116 4.7 815 12 US-10-335-977-7411 Sequence 7411, Ap

33 116 4.7 842 12 US-10-335-977-7412 Sequence 7412, Ap

34 115.5 4.7 530 12 US-10-282-122A-48561 Sequence 48561, A

35 115.5 4.7 959 16 US-10-616-624-12 Sequence 12, Appl

36 115 4.7 292 12 US-10-282-122A-58541 Sequence 58541, A

37 115 4.7 518 12 US-10-282-122A-44255 Sequence 44255, A

38 114 4.6 290 9 US-09-815-242-13182 Sequence 13182, A

39 114 4.6 290 12 US-10-282-122A-74177 Sequence 74177, A

40 113.5 4.6 515 9 US-09-486-734A-6 Sequence 6, Appli

41 113.5 4.6 1026 15 US-10-369-493-3277 Sequence 3277, Ap

42 113 4.6 317 12 US-10-282-122A-74345 Sequence 74345, A

43 113 4.6 518 9 US-09-815-242-12279 Sequence 12279, A

44 113 4.6 518 12 US-10-282-122A-44253 Sequence 44253, A

45 113 4.6 518 12 US-10-282-122A-70377 Sequence 70377, A

## ALIGNMENTS

### RESULT 1

US-09-966-997-11  
; Sequence 11, Application US/09966997  
; Publication No. US20030100052A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Shuang-yong  
; APPLICANT: Dore, Andrew  
; APPLICANT: Hume, Adam  
; APPLICANT: Pelletier, John  
; APPLICANT: Zhou, Jing  
; TITLE OF INVENTION: Method For Cloning And Expression of BsmBI Restriction  
; TITLE OF INVENTION: Endonuclease And BsmBI Methylase In E. coli And  
; TITLE OF INVENTION: Purification Of BsmBI Endonuclease  
; FILE REFERENCE: NEB-191  
; CURRENT APPLICATION NUMBER: US/09/966,997  
; CURRENT FILING DATE: 2001-09-28  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 11  
; LENGTH: 1068  
; TYPE: PRT  
; ORGANISM: Bacillus stearothermophilus B61  
US-09-966-997-11

Query Match 9.0%; Score 223; DB 10; Length 1068;  
Best Local Similarity 20.7%; Pred. No. 1.9e-12;  
Matches 112; Conservative 86; Mismatches 219; Indels 124; Gaps 21;

QY 21 TTAQVLQGFPTPKMAATLMSMLRVDDLRGTVRVDPFGAGVGSLLTA-ALVDRLHTRPDV 79  
118 TASKAVGYFTPTSLGTRVMVKLAKDKPKNLKSIIVDPACGIGSLALALI-----YNPEI 172  
QY 80 AVHVAVETDFPVVPLRATLEECRNAYGIS--YDLVEG---DYLLNQAKLDPGLDVI 134  
173 E-NVWGIELDSFTANISHKLLVRISKDLGTPKIKINQNFLLVNLVYEEHKEFDLLI 231  
QY 135 ANPPYCKL-----ASDSLRLATTARAVDVP 160

Db 232 MNPYGRVRFLKNSLTNETKSGTLEGI SELEKLRREETILMAADLRKKFASVGLKGT 291  
Qy 161 NVYVAFWRAVISLKEQGRGVFIIVPRSWANGPYQPRHMLMTAVSLDLHVFESRTKVF 220  
Db 292 EYSKFLAISTKIVKONGVIAITPSSWLGDESGRELKYLVENHGISCINWFKESAKLIF 351  
Qy 221 ADTKQENVIIVAFSVRQSSVWLSRSVAHGEESIASVSPFSAVHDEDDDKIVHFAES 280  
Db 352 SG--VNQPTTVKIKVNESKIEIQ-----PLSSL---EELGRDQYLD 393  
Qy 281 ASVP--SAARFTLADLGIG-----VSTGKVDPRNQYITDNDASGVVPM 324  
Db 394 CNIKKYSPEWYRIPOCGNERAKILSKLHNHAPLSHKKIYNLRGELDTLTHKDLLSDNPN 453  
Qy 325 VYQSNIRSGKIDWPVGARKPO-----GFVAVEDVAL-----RQLIPO 362  
Db 454 HWRL--IRGDHVE--KFNLPNPESEKLGIV---DHQLFKRMGKSNKLRHKNWRITLPO 507  
Qy 363 GSYVVKQTAKEDERRRVIAAVWDGASRVALDNKTNYLH--ESQPLEKNVARGLMWLNS 421  
Db 508 CSYM-----NKKRIEACIVEPNNIIA--NSCNYITLEDGELVONLL--LCAINS 556  
Qy 422 TVLDQYFRAFSGHTOVNAGDLRLPFLC-----REDLILAKVVPDGLPDQETLDVAVRL 477  
Db 557 AVIEWRFLFNSNNHVSVEIDEFFIFKFDTEMTMLKGLFLKPIENWSKIEALIALM 616  
Qy 478 F 478  
Db 617 Y 617

## RESULT 2

US-10-335-977-7269  
; Sequence 7269, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7269:

SEQUENCE CHARACTERISTICS:

LENGTH: 543 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

## ORIGINAL SOURCE:

ORGANISM: Helicobacter pylori

## FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...543

SEQUENCE DESCRIPTION: SEQ ID NO: 7269:

US-10-335-977-7269

Query Match 6.6%; Score 164; DB 12; Length 543;

Best Local Similarity 23.3%; Pred No. 5.2e-07;

Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;

Qy 28 QFFPMKATLMASRLRVDDLRGTVRVLDPGAGVSLTAALVDRJLHTRPDPVAVHVAVE 87

Db 208 EYVTELSIASIIAKLL-VNKPTQSVKIYDPSAGTGTLLMALAHQIGTD--SCTLYAQDIS 264

Qy 88 TDPFVVPYLRATLECRNAYGISYDL---VEGDYLLNQAKLD--GPFDLVIANPYGKL 142

Db 265 QKSLRMLKLIILND-----LTHSLRHAIEGNTLTNPYHSHKDHKGKMDFIYSNPPFKLD 318

Qy 143 ASDSLARLATTAR--AVDPVN-----VYVAFWRAVISLKEQGRGVFIVPR---S 187

Db 319 FSNEHAEISQNKDNDFLGVNIPKNDKSKMPLYTLFHOCLNMLSPKKGALIVTGFIS 378

Qy 188 WANGPYRQFRHMLTAVSLDILHVFESRTKVFAADTKVKQENV-IVAFSVRQSSSVLS 246

Db 379 AKSGVNNKNNRHLVDERL---VYGVICMPSQVFANTGT---NVSIIFFOKTSPSAKEVILI 432

Qy 247 RSVANGEE-----SIASSVPFSALV-HDEDDDKIVHFAESA 281

Db 433 DASKLGEETENKNNKTRLPDSMDMLILETFONKAPKSDFCALVSFDEITEK-----NY 486

Qy 282 SVPSAARFTLADLGIGVSTGKVVDPRN--ROYLTD 314

Db 487 SLNPGQYTIETDSETISQ---AEFENLMQYSSSE 518

## RESULT 3

US-10-335-977-7268

; Sequence 7268, Application US/10335977

; Publication No. US20040052799A1

; GENERAL INFORMATION:

APPLICANT: DOUGLAS SMITH et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
RELATING TO HELICOBACTER PYLORI FOR  
DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031

CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD

STREET: 28 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109-1875

COMPUTER READABLE FORM:

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: Windows NT 4.0

SOFTWARE: UNIX

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/335,977

FILING DATE: 30-Dec-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/993,002

FILING DATE: 17-DEC-1997

ATTORNEY/AGENT INFORMATION:

NAME: Mandragouras, Amy E.

REGISTRATION NUMBER: 36,207

REFERENCE/DOCKET NUMBER: GTN-018

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617)227-7400

TELEFAX: (617)742-4214

INFORMATION FOR SEQ ID NO: 7268:

```
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 543 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: YES
;   ORIGINAL SOURCE:
;     ORGANISM: Helicobacter pylori
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...543
;   SEQUENCE DESCRIPTION: SEQ ID NO: 7268:
US-10-335-977-7268

Query Match          6.5%; Score 161; DB 12; Length 543;
Best Local Similarity 23.3%; Pred. No. 1e-06;
Matches 78; Conservative 61; Mismatches 124; Indels 72; Gaps 16;

QY 28 QFTFPMKAATLMASMLRVDDLRTGTVRLDPGAGVGSILTAALVDRLHTRPDVAHVHVAVE 87
Db
QY 208 EYITPLSIASIIAKLL-VNKPTQSVKIYDPSAGTGLMALAHQIGTD--SCTLYAQDIS 264
Db
QY 88 TDFVVPYLRATLEECRNAYGISYDL---VEGYLLNQAKLD--GPFDLVIANPPYKGL 142
Db
QY 265 QKSLRMLKLNLIIND-----LTHSLRHAIEGNTLTNPYHSKDHKGKMBFIVSNPPFKLD 318
QY 143 ASDSLARLATTAR--AVDPN-----VYVAFWRRAVISLKEQGRGVFIVPR---S 187
Db
QY 319 FSNHEAEISQNKNDFFLGVNIPKNDKSKMPIYTLFFQHCLNMLSPKKGAIIVPTGFTS 378
QY 188 WANGPYVROFRHMLTAVSLDILHVFESRTKVPADTKVKQENV-IVAFSVRPOSSSVWLS 246
Db
QY 379 AKSGVENKIVRHLDVRL---VYGVICMPSQVPANTGT---NVSIIFQKTPSAKEVILI 432
QY 247 RSVAHGEE-----SIASSVPFSALV-HDEDDDKTVHFAESA 281
Db
QY 433 DASKLGEETENKNNKTRLRPSMDLILETFONKAPKSDFCALVSFDEITEK-----NY 486
QY 282 SVPSAARFTLADIGIGVSTGKVVDFRN--ROYLTD 314
Db
QY 487 SLNPGQVFTIEDTSETISQ---AEFENLMQOYSSE 518

RESULT 4
US-10-282-122A-51640
; Sequence 51640, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 543 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: YES
;   ORIGINAL SOURCE:
;     ORGANISM: Helicobacter pylori
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...543
;   SEQUENCE DESCRIPTION: SEQ ID NO: 7268:
US-10-335-977-7268

Query Match          5.7%; Score 141; DB 12; Length 571;
Best Local Similarity 19.7%; Pred. No. 0.00011;
Matches 112; Conservative 85; Mismatches 213; Indels 158; Gaps 27;

QY 19 DPTQAVLQGFPTPMKAATLMASMLRVDDLRTGTVRLDPGAGVGSILTAALVDRLHTRPD 78
Db
QY 6 DATERLTGRYFTTPDLASYIIDWIQDN--NINKILEPSCGNGVFECLGER----RLE 59
QY 79 VAVHVVAVTDFVVPYLRATLE-----ECRNAYGISYDLVEGYLLNOG----- 123
Db
QY 60 DGRNITAIBEDDV--SFEASMQIDNSLRFNFCYQALRNDGDIVNNGIVINDDFYK 117
QY 124 ---AKLDGP-FDLVIANPPY-----GKLASDSLARLATTARAVDVNVPVAVFVRAV 171
Db
QY 118 VTEQLQGRFOAIVGNPPYRYQYLSQEQREBSKILIRNNRPNKILINAWFSFVACA 177
QY 172 ISLKEQGRGVFIVPRSWANGPYRQFRHMLTAVSLDILHVFESRTKVPADTKVKQENVI 231
Db
QY 178 EILDGNGKMLVIPAEELLQVAYAEIDLRRFTMRFTQITIVTF--RELVPFN--VQEEVL 233
QY 232 V-----AFSVRPOSSSVWLSRVAH--GEESIASVSPFSALVHDE----- 269
Db
QY 234 LLVEKEILHTRHQRLRIVEYQDINELTESNDLDEYFPFNDVEINESKWTYFLSANDIRLI 293
QY 270 ----DDDKLVHFAESASVPSAARFTLADLIGVSTGKVVDFRNROYLTONTLDASGVV-PM 324
Db
QY 294 NNIRENDKFRFSDVARVE-----VGITGNNDYFCVNRVRVEYDILERVCRPL 342
QY 325 V-----YQSNIRSGK-----IDWPQVGARKPQGFVAVEDVALRQLLPQGSY 365
Db
QY 343 IARSVNNGVVKFTYEDWKSNIENGAKTYLIDFPDVPYDEVQ-----LSYKQVIEYGIH 395
QY 366 VVVKROTAK---EDRRRVIAAVW-----DCASRVALD 394
Db
QY 396 -NDKNSGYKCRIDRWYKVPESIWVPDAFFLRRNHLYPKFLVINSQEVNAVSTDMHRVFN 454
QY 395 NKTNYLHESQRPLEKNVARGCLMLNLTVDQYFRAFSGHTQVNAAGDLRLPFLCREDLI 454
Db
QY 455 N-----HDDRE-----RILSYNSIAL--AFTEIGRS--YGGGVLEILP---RE--- 493
QY 455 LLAKVVPDGLPDQETL-DAVVARLFCEI 481
Db
QY 494 -VGRIMPDLRRELIDDATVSNLLNRI 520

RESULT 5
US-10-282-122A-65227
; Sequence 65227, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65227
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-282-122A-65227

Query Match      5.5%; Score 135; DB 12; Length 514;
Best Local Similarity 22.0%; Pred. No. 0.00038;
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;

Qy 27 QGFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGSLSAALVDRHLHTERPDVAHVAV 86
Db 196 GEFTPOSVSKLIA-LRAVHGOEKVKNKYDPACGSGSL-----LQAKKQFDEHI--I 245
Qy 87 ETDPPVVPYLRATLEECRN---AYGISYD---LVEGDYLLNOGAKLDGPFDLVIANPPYG 140
Db 246 BEGFGGOEINHTTYNARMNFMFLHNVNKHEIELGDTLTNPKLKDSPFFDAVSNPPYS 305
Qy 141 -----KLASDLARLATTARAVDPNVYVAFWVRVAVISLKEQGRGVFI-VPRSWAN 190
Db 306 IDWIGSDPTLINDD--RFAPAGVLAPSKADFAFILHALNVLSGRGAIVSPFGIFYR 363
Qy 191 GPYRQRFHWM-----TAVSLDILHVFESEKTKVPFADTKVQENVIV 232
Db 364 GGAQKIROYLVEGNYVETVIALAPNLPFYGCIANVILVLSKHK-----DNTDI 412
Qy 233 AF-----SVRPQSSSVLSRSVAHGEESIASVPFSALVHDEDDDKIVHPAESASVP-- 284
Db 413 QFIDAGGFFKETNNVLT-----BEHIAEIVKL-----FADKADVPHIA 452
Qy 285 -SAARFTLADIGVSTGKVDFFNRQYLT-----NLDASGVPMV 325
Db 453 QNAAQQTVDKNGYNLAVSYVEAETREVIDIRQINAEISVTVAKI 498

RESULT 6
US-10-282-122A-65854
```

```
; Sequence 65854, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65854
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-10-282-122A-65854

Query Match      5.4%; Score 134; DB 12; Length 514;
Best Local Similarity 22.0%; Pred. No. 0.00048;
Matches 76; Conservative 52; Mismatches 128; Indels 90; Gaps 15;

Qy 27 QGFTPMKAATLMASMLRVDDLRTGTVRVLDPGAGVGSLSAALVDRHLHTERPDVAHVAV 86
Db 196 GEFTPOSVSKLIA-LRAVHGOEKVKNKYDPACGSGSL-----LQAKKQFDEHI--I 245
Qy 87 ETDPPVVPYLRATLEECRN---AYGISYD---LVEGDYLLNOGAKLDGPFDLVIANPPYG 140
Db 246 BEGFGGOEINHTTYNARMNFMFLHNVNKHEIELGDTLTNPKLKDSPFFDAVSNPPYS 305
Qy 141 -----KLASDLARLATTARAVDPNVYVAFWVRVAVISLKEQGRGVFI-VPRSWAN 190
Db 306 INWIGSDPTLINDD--RFAPAGVLAPSKADFAFILHALNVLSGRGAIVSPFGIFYR 363
Qy 191 GPYRQRFHWM-----TAVSLDILHVFESEKTKVPFADTKVQENVIV 232
Db 364 GGAQKIROYLVEGNYVETVIALAPNLPFYGCIANVILVLSKHK-----DNTDI 412
Qy 233 AF-----SVRPQSSSVLSRSVAHGEESIASVPFSALVHDEDDDKIVHPAESASVP-- 284
Db 413 QFIDAGGFFKETNNVLT-----BEHIAEIVKL-----FADKADVPHIA 452
Qy 285 -SAARFTLADIGVSTGKVDFFNRQYLT-----NLDASGVPMV 325
```

Db 453 QNAACQTVKDNKNGYNLAVSSVVEPEDTREIIDIKOLNAEISETVAKI 498

## RESULT 7

```

US-10-282-122A-72313
; Sequence 72313, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72313
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72313

```

Query Match	5.4%	Score 134;	DB 12;	Length 534;
Best Local Similarity	20.9%	Pred. No. 0.00051;		
Matches	62;	Conservative 56;	Mismatches 147;	Indels 32; Gaps 11;
Qy	27	GQFFTPMKAATLMASMLRVDDLRCTVGLDPGAGVGSITAAVLDRLLHTERPDVAHVAVV	86	
Db	203	GEFFTPQTVSKLLAQLVMVGK-EHINKVYDPTCGSGSLLLOKMKQFETHILEEFGFOEI	261	
Qy	87	ETDPFVVPYLRATLEBCRNAGVISDLVEGDPYLLNQCAKLDGPDFDLVIANPPYG----	140	
Db	262	NMTNYNLARMNMFEL---HNINYNFDIRRGDTLLNPQHLVPERPDAIVSNPPYSIKWIGD	318	
Qy	141	---KLSDSLARLATTARAVDVNPVYVAFWRAVISLKEGCGVFI-VPRSWANGPPYRQ	196	
Db	319	ADPTLLINDE--RFAPAGKLAPSKADTAFIMHSLUSHLNKGRAAIVCFPGGIFYRGGAET	376	
Qy	197	FRHLWMTAVSLDILHVFESRTKVFADTKVQENIVAFSPQSSSVVL-----SRSVAH	251	
Db	377	IROYLIDNNFVEA--VIALPNLFVGTSIATYIVLAKN-KPEDKTAFIDASSDEKSTVS	433	

**Qy**      252 GEESIASVPFSAVHDEDDKIVH-FAESASVPSAARFT----LA---DLGIGVST 300  
          | :         :: :: :: | | | | : | : | : |  
**Dp**      434 GKNGEYETVINGKIINPKNIEALVEIFKNKKDQVDYEAKLVDNKLIAENDYNLSVST 490

## RESULT 8

```

US-10-335-977-7481
; Sequence 7481, Application US/10335977
; Publication No. US20040052799A1
; GENERAL INFORMATION:
; APPLICANT: DOUGLAS SMITH et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
; RELATING TO HELICOBACTER PYLORI FOR
; DIAGNOSTICS AND THERAPEUTICS
;
; NUMBER OF SEQUENCES: 10031
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: UNIX
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/335,977
; FILING DATE: 30-Dec-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/993,002
; FILING DATE: 17-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: GTN-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
;
; INFORMATION FOR SEQ ID NO: 7481:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Helicobacter pylori
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...381
; SEQUENCE DESCRIPTION: SEQ ID NO: 7481:
;
US-10-335-977-7481
Query Match          5.4%; Score 133.5; DB 12; Length 381;
Best Local Similarity 26.6%; Pred.No. 0.00033;
Matches 54; Conservative 29; Mismatches 73; Indels 47; Gaps 10;

QY      26 LGQFFTPMKAATLMASMLRVDDLGRVVLDPGAGVGLTAALVDRLHTERPDVAHVHA 85
Db      13 LGQVFPT--KNIVDFMLIIKHNHGS--VLEPSAGDSFLKRL-----KKAVG 55
QY      86 VETDPPFVPVYLRTLEEBCRNAYGISYDLVEGDYLLNQAKLDGPFDLVIANPPYK---L 142
Db      56 IEIDPKICP-----KNALCMDF----FDY-----PLENQFTLIIGNFPVVKHKDI 96
QY      143 ASDSLARLATARAQDVNVVAFWFAVLSLKEQGRGVFIIVPRSWANGPYRQFRHWM 202
Db      97 APSTKEKLHYSL-FDERSSYLLFIEIKAIKHLFKPGELIIFITPRDFLKSTSSVKLNEMIW 155
OV      203 TAVSLDILIHVFESRTKVFAADTKV 225

```

Db 156 KEGT--ITHFFE-----LGDQKI 171

RESULT 9

US-10-335-977-7482  
; Sequence 7482, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR  
; DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-DEC-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7482:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 384 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...384  
SEQUENCE DESCRIPTION: SEQ ID NO: 7482:  
US-10-335-977-7482

Query Match 5.4%; Score 133.5; DB 12; Length 384;  
Best Local Similarity 26.6%; Pred. No. 0.00034;  
Matches 54; Conservative 29; Mismatches 73; Indels 47; Gaps 10;

QY 26 LGQFTPMKAATLMASMLRVDDLRTGTVRLDPGAGVSLTAALVDRLHTERPDVAHVVA 85  
Db 16 LGQVTFP-----KNIVDFMLTKLNHGS--VLEPSAGDGSFLKRL-----KXAVG 58  
QY 86 VETDPFVVPYLRATLEECNAYGISVDLVEGYYLLNQAKLDGPFDLVIANPPYCK--L 142  
Db 59 LEIDPKICP-----KNALCMDF----FDY-----PLENQFTIIGNPPYVKHKDI 99  
QY 143 ASDSLARLATTARADVDPNVYAFWVRVAVISLKEQGRGVFIYVRSWANGPYVRQPRHMLM 202  
Db 100 APSTKEKLHSL-FBERSNLFFTEKAIKHLKPKGELIFITPRDFLKSSTSSVKLNWIY 158  
QY 203 TAVSLDILHVFESRTKVPADTKV 225  
Db 159 KEGT--ITHFFE-----LGDQKI 174

RESULT 10

US-09-882-227-600  
; Sequence 600, Application US/09882227  
; Publication No. US20030158396A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Oomenen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; RELATING TO HELICOBACTER PYLORI

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/882,227  
FILING DATE: 30-DEC-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragoras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7482:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 384 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...384  
SEQUENCE DESCRIPTION: SEQ ID NO: 7482:  
US-10-335-977-7482

Query Match 5.3%; Score 130; DB 10; Length 545;  
Best Local Similarity 18.9%; Pred. No. 0.0013;  
Matches 86; Conservative 87; Mismatches 203; Indels 78; Gaps 19;

QY 17 KLDPTTQAVLGQFTPMKAATLMASMLRVDDLRTGTVRLDPGAGVSLTAALVDRLHTER 76  
Db 101 ELSNATRNLEGIYTPNRIVEQLFTLPKDFVDSQAI-FCDDPAVSGNFI-----WHALK 153  
QY 77 PDVAV-HVVAVETDPFVVPYLRATLEECNAYGISVDLVEGYYLLNQAKLDGPF 130  
Db 154 LGFVENIYGYDTDAFAVALTKKRIKERYHLDCLN-----IVQKDFL---NLKHTPOF 203  
QY 131 DLVIANPPYCKL-----ASDSLARLATTARADVDPNVYAFWVRVAVISLKEQGRGVFIYVR 186  
Db 204 DCIFTNPPWPKYNNQKNENKQFNLSQSLDSASL---FFIASLNCLENKNAHLGLLPE 260  
QY 187 SWANGPYVRQPRHMLTAVSLDILHVFESRTKVPADTKVQK---ENVIVAFVSRP---QS 240  
Db 261 SCLNIDAFKKQREMAK-----FHIRSLIDFKPKNLTAKVGLAKKTKNKOQK 311  
QY 241 SSVLSRSVAHGEESIASVSP---FSALVHDEDDDKIVHFAESASVPSAARFTLADLGIG 297  
Db 312 ISCFYQNSKFKRSPSSFFNNPKKIFNIHCSSKENKILDHL---FSLPHMTLKNNAHFALG 368  
QY 298 VSTGKVVDFRQRYLTDLNLSAGVVPVYQNSIRSGKIDWPOVGARKPGQFVAVEDVALR 357  
Db 369 IVTG-----NNKEKLHPKQE-KNTIPIFRGSDILK-----GLKAPSQFTNAGLKDCQ 415  
QY 358 QLLPGQSY-----VVVKQTKAEDRRRVIAAVWDGASRVALDNKTNLYHESORPLEKXVA 412  
Db 416 QVAPLSLQAREKIVYKFISSK-----LVFFYDNKQRLFLNSANFVYKFNFINAHAL 469  
QY 413 RGLMLMLNSTVLVDQYFRAPSGHTQVNAAGDLRLRP 446  
Db 470 KEL---LNSDLMQFTFESLFKTHKILRKDLECLP 500

RESULT 11

US-10-335-977-7271  
; Sequence 7271, Application US/10335977  
; Publication No. US20040052799A1  
; GENERAL INFORMATION:  
; APPLICANT: DOUGLAS SMITH et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; RELATING TO HELICOBACTER PYLORI FOR

## DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 10031  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: Windows NT 4.0  
SOFTWARE: UNIX  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/335,977  
FILING DATE: 30-Dec-2002  
PRIOR APPLICATION NUMBER: 08/993,002  
FILING DATE: 17-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: GTN-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 7271:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 545 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (B) LOCATION 1...545  
SEQUENCE DESCRIPTION: SEQ ID NO: 7271:  
US-10-335-977-7271

Query Match 5.1%; Score 126; DB 12; Length 545;  
Best Local Similarity 17.5%; Pred. No. 0.0034;  
Matches 83; Conservative 96; Mismatches 185; Indels 110; Gaps 21;  
QY 17 KLDPTTQAVLGQFPTPKAATLMASMLRVDLGRGTVRVLDPGAGVGSLSLAALVDRLHTR 76  
DB 101 ELNTRNLEGIYTPNRIVEQLFTLPKDFTTQAI-PCDPVAVSGSNFV-----MHALK 153  
QY 77 PDVAV-HVAVETDPFVVPYLRATLEE-----CRNAYGISYDLVEGDYLLNQAGKLDGPF 130  
DB 154 LGPKVENIYGYDTDAFAVALTKKRIKERYRLDCFN-----IMQKDFL---SLKHAPQF 203  
QY 131 DIVIANPPYKLL-----ASDSLRLATTARAVDPNVVAVFVRAVISLSKQGRGVETVPR 186  
DB 204 DCIFTNPWGKKYNQONKFNKQFNHQLSDLSASL---PFMASLNCLEKNAYLGLLLE 260  
QY 187 SWANGPYV-----RQFRHMLMTAVSILDLHVFSRTKV-----FADTK 224  
DB 261 SCLNIDAFSKMREKALFQMRSLDFDKPFKMLTKAVGLKALKTHNKDKQKISCFYQNSE 320  
QY 225 VKQENVIVAFSPQSSSVLSRVAHEGESIAS---SVPSALVHDEDDDKIVHPAES 281  
DB 321 FKRSF--SSFPNPK---IPNHCESHQKNKLDHLSPLHTLKN-----AHFA--- 366  
QY 282 SVPSAARFTLADLIGVSTGKVDPRKQYLTDLNDSGVVPMYVQNSIRSGKIDWQVG 341  
DB 367 -----LGIWVG-----NNKEKLHPKQE-KNTIPFRGSDILKDRL----- 400  
QY 342 ARKPGFVAVEDVALROLLPQGSY-----VVVKQTAKEDRRRVIAAVWDCASRVLDNK 396  
DB 401 -KAPSQFINADLKCCQVAPLSLQAREKIVYKFISK-----LVFFYDNKQRLFLNSA 453

QY 397 TNYLHESQRFLEKNVARGMLMLNSTVLDQYFRAFSGHTQVNVAGDLRRPLFLCR 450  
DB 454 NMFLKENFPINAHALKEL---LNSDLMQFIFESLFKTHKILRKDLCECLPLFAQ 504  
RESULT 12  
US-09-712-363-185  
Sequence 185 Application US/09712363  
Patent No. US20020164588A1  
GENERAL INFORMATION:  
APPLICANT: Eisenberg, David  
APPLICANT: Rotstein, Sergio H.  
APPLICANT: Marcotte, Edward M.  
TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
FILE REFERENCE: 07419-032001  
CURRENT APPLICATION NUMBER: US/09/712,363  
CURRENT FILING DATE: 2000-11-13  
PRIOR APPLICATION NUMBER: PCT/US00/02246  
PRIOR FILING DATE: 2000-01-28  
PRIOR APPLICATION NUMBER: 60/179,531  
PRIOR FILING DATE: 2000-02-01  
PRIOR APPLICATION NUMBER: 60/117,844  
PRIOR FILING DATE: 1999-01-29  
PRIOR APPLICATION NUMBER: 60/118,206,  
PRIOR FILING DATE: 1999-02-01  
PRIOR APPLICATION NUMBER: 60/126,593  
PRIOR FILING DATE: 1999-03-26  
PRIOR APPLICATION NUMBER: 60/134,093  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/134,092  
PRIOR FILING DATE: 1999-05-14  
PRIOR APPLICATION NUMBER: 60/165,124  
PRIOR FILING DATE: 1999-11-12  
PRIOR APPLICATION NUMBER: 60/165,086  
PRIOR FILING DATE: 1999-11-12  
NUMBER OF SEQ ID NOS: 292  
SOFTWARE: PastSeq for Windows Version 4.0  
SEQ ID NO 185  
LENGTH: 317  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-09-712-363-185  
Query Match 5.1%; Score 125.5; DB 9; Length 317;  
Best Local Similarity 25.6%; Pred. No. 0.0016;  
Matches 73; Conservative 48; Mismatches 109; Indels 55; Gaps 13;  
QY 4 IASTETTRQAAIGKLDPTTQAVLGQFPTPKAATLMASMLRVDLGRGTVRVLDPGAGVGS 63  
DB 14 LGTEIRRLAK--ELDFRPRKSLGQNF--VHDANTVRRVVAASGVSRSDLVLEVGPLGS 69  
QY 64 LTAALVDRLHTRPDVAVHVAVETDPFVVPYLRATLEECRNAYGISYDLVEGDYLLNQ 123  
DB 70 LTLALLDRGAT-----VTAVEIDPLLASRLQQTVAEHSSEVHRLTVNRDVLALRR 121  
QY 124 AKLDGPFDLVIANPPYKGLASDSLRLATTARAVDPNV-VYAFVWPAVLSK---BQGR 179  
DB 122 EDLAAAPTAVVNLPLY-NVAVPALLHL-----LVEPESIRVVTVMQAEVRLAAEPGS 175  
QY 180 GVFTVPSRWANGPYKQFRH-----WLMTAVSLDLHV--FES-----RTKV 219  
DB 176 KEYGVPSS--VKLRFGRVRCGMVSPVTFPIRVVSGLVIRIDRYETSPWETDDAFRRV 233  
QY 220 FADTKVKQENVIVAFSPQSSSVLSRVAHEGES-----IASV 260  
DB 234 F-----ELVDIAFAQRRTSRNAPVQWAGSGSESANRLAASI 271  
RESULT 13  
US-10-282-122A-62340  
Sequence 62340, Application US/10282122A

Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 62340

LENGTH: 317

TYPE: PRT

ORGANISM: Mycobacterium bovis

US-10-282-122A-62340

Query Match 5.1%; Score 125.5; DB 12; Length 317;  
Best Local Similarity 25.6%; Pred. No. 0.0016;  
Matches 73; Conservative 48; Mismatches 109; Indels 55; Gaps 13;  
Qy 4 IASTETRRQALGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLGRGTVRVLDPGAGVGS 63  
Db 14 LGREIRRLAK--ELDFPRKSLQNF--VHDANTVRVVAASGVSRSDLVLEVPGLGS 69  
Qy 64 LTAALVDRLHTRDPDVAVHVVAVETDPFVVPYLPATLEECRNAYGISYDLVEG DYLLNQ 123  
Db 70 LTLALLDRGAT-----VTAVEIDPLLASRLQQTVAEHSSEVHRLTVVNRDVLALRR 121  
Qy 124 AKLDGPPDLVIANPPYKGLASDSLRLATLATTARAVDENV-VYAFWVRAVISLK---BQGR 179  
Db 122 EDLAAATAVANLUPY-NVAVPALLHL-----LVEFPIRVVTVMQAEVRLAEPGS 175  
Qy 180 GVFTVPRSWANGPYRQFRH-----WMTAVSLDILHV--FES-----RTKV 219  
Db 176 KEYGVPS--VKLRFGRVRCGMVSPVTFVPIRVYSGLVRIIDRYETSPWPTDDAFRRV 233  
Qy 220 FADTKVQENVIYAFSVRPOSSSVLSRSVAHGEES-----IASSV 260  
Db 234 F-----ELVDIAFAQRKTSRNAFVQWAGSGSESANRLAASI 271

RESULT 14

US-10-282-122A-64484  
Sequence 64484, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari  
APPLICANT: Zyskind, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 64484

LENGTH: 317

TYPE: PRT

ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-64484

Query Match 5.1%; Score 125.5; DB 12; Length 317;  
Best Local Similarity 25.6%; Pred. No. 0.0016;  
Matches 73; Conservative 48; Mismatches 109; Indels 55; Gaps 13;  
Qy 4 IASTETRRQALGKLDPTTQAVLGQFFTPMKAATLMASMLRVDDLGRGTVRVLDPGAGVGS 63  
Db 14 LGREIRRLAK--ELDFPRKSLQNF--VHDANTVRVVAASGVSRSDLVLEVPGLGS 69  
Qy 64 LTAALVDRLHTRDPDVAVHVVAVETDPFVVPYLPATLEECRNAYGISYDLVEG DYLLNQ 123  
Db 70 LTLALLDRGAT-----VTAVEIDPLLASRLQQTVAEHSSEVHRLTVVNRDVLALRR 121  
Qy 124 AKLDGPPDLVIANPPYKGLASDSLRLATLATTARAVDENV-VYAFWVRAVISLK---BQGR 179  
Db 122 EDLAAATAVANLUPY-NVAVPALLHL-----LVEFPIRVVTVMQAEVRLAEPGS 175  
Qy 180 GVFTVPRSWANGPYRQFRH-----WMTAVSLDILHV--FES-----RTKV 219  
Db 176 KEYGVPS--VKLRFGRVRCGMVSPVTFVPIRVYSGLVRIIDRYETSPWPTDDAFRRV 233  
Qy 220 FADTKVQENVIYAFSVRPOSSSVLSRSVAHGEES-----IASSV 260  
Db 234 F-----ELVDIAFAQRKTSRNAFVQWAGSGSESANRLAASI 271

Job time : 105.928 secs

```
RESULT 15
US-10-282-122A-53706
; Sequence 53706, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53706
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Corynebacterium diptheriae
US-10-282-122A-53706

Query Match          5.0%; Score 124.5; DB 12; Length 295;
Best Local Similarity 27.9%; Pred. No. 0.0018;
Matches 57; Conservative 25; Mismatches 81; Indels 41; Gaps 8;

QY      4  IASTETRRQALGKLDPTTQAVIGQFTTPMKAATLMASMLRVDDLRCGTVTVLDPGAGVGS 63
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      10  LGEVEIRQLAE--KLDVTPTKLGQNF--VHDPNTVMIVSAADLNSDDHVEVGPGLGS 65

QY      64  LTAALVDRLHTERPDVAHVAVVETDPFVVPVPLRATLEECRNAYGISYDLVGEDYLLNQG 123
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      66  LTLALL-----DTAQKVAVEIDPRLAQQLPLTVAEERAGQPADRLNIHKDALTVP 117

QY      124  AKLDGPFDLVIANPPYG-----KLASDSLARIATTA--RAVDVP 160
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      118  DDIDHPTALV-ANLPYNVSVPLVLLHLIQIFPTIRRLVMVQAEVADRLLAADPGNKVYGP 176

QY      161  NVYVAFW--VRAVISLKEQGRGVF 182
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      177  SVKASFYGNVRRAGSI---GKNVF 197
```

Search completed: October 1, 2004, 16:28:45

**This Page Blank (uspto)**

Result	Query No.	Score	Match	Length	DB	ID	Description
C	1	39.6	4.1	1201	13	EX381961	EX381961 EX381961
C	2	39.2	4.0	1201	9	AL544900	AL544900 AL544900
C	3	38.8	4.0	808	14	CK126515	CK126515 AGENCOURT
C	4	38.6	4.0	539	9	AA788783	AA788783 ah31b05.s



High quality sequence stop: 690.

## FEATURES

Location/Qualifiers  
1. .808  
/organism="Danio rerio"  
/mol\_type="mRNA"  
/db\_xref="taxon:7955"  
/clone="IMAGE:7049194"  
/tissue\_type="whole body"  
/lab\_host="DH10B"  
/clone\_lib="NIH\_ZGC\_10"  
/note="Vector: pExpress1; Site 1: NotI; Site 2: EcoRV;  
Bulk tissue was collected from a whole adult individual  
from the Tuebingen strain. 1st strand cDNA was primed with  
a Not I - oligo(dT) primer, double-stranded cDNA was  
cloned into the Not I and EcoRV sites of pExpress-1.  
Library was size-selected for >1 kb fragments. A  
normalized version of this library is also available  
(NIH\_ZGC\_7). Library was constructed by Open Biosystems  
(Huntsville, AL)."

## ORIGIN

Query Match 4.0%; Score 38.8; DB 14; Length 808;  
Best Local Similarity 53.7%; Pred. No. 8.8;  
Matches 79; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 48 TCGCGGGCGTCTAAAGCGTTTGGTTTACGGCGCAACGATATAGCTCGTAGCGC 107  
Db 770 TCGGAGTGGCATCTTCATGCAGCTCGCGTTTCAGCTTTGGCGATGAGCTCCAGGTGGCA 711  
QY 108 TGTGACATTCCTCGCGCTTTCGCGTTTGAAGCGGGAGATCGTGGTTGACTCCGACAC 167  
Db 710 CDTGTCCCTGACAGCTCGATTCACTTTGAAGTCTGCTGATCATCGGTTGAGATCAGCAA 651  
QY 168 TCCAGCGCTTGGCGTTTCAGAGATCAT 194  
Db 650 TCTCAGCCTTGGTTTACGAGGTGAT 624

## RESULT 4

AA788783

LOCUS

DEFINITION

ab31b05.s1 Soares parathyroid tumor NBHPA Homo sapiens cDNA clone  
1240401 3', similar to gb:K01144 HLA CLASS II HISTOCOMPATIBILITY  
ANTIGEN, GAMMA CHAIN PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION

AA788783

VERSION

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 539)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgaps-r@mail.nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bbrp/image/image.html  
Insert Length: 1264 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 285.  
Location/Qualifiers  
1. .539  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

## FEATURES

source

/clone="1240401"  
/tissue\_type="parathyroid tumor"  
/dev stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/clone\_lib="Soares parathyroid tumor NBHPA"  
/note="Organ: parathyroid gland; Vector: pT7T3D  
(Pharmacia) with a modified polylinker; Site 1: Not I;  
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
oligo(dT) primer  
(5'-TGTTACCAATCTGAAGTGGGCGGCACCAATTTTTTTTTTTTTTTT  
TTTTT-3'), double-stranded cDNA was size selected, ligated  
to Eco RI adapters (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of a modified pT7T3  
vector (Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by Bento  
Soares and M Fatima Bonaldo. RNA from sporadic parathyroid  
adenomas was kindly provided by Dr. Stephen Marx, National  
Institute of Diabetes and Digestive and Kidney Diseases,  
NIH."

## ORIGIN

Query Match 4.0%; Score 38.6; DB 9; Length 539;  
Best Local Similarity 50.8%; Pred. No. 8;  
Matches 92; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 709 CATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGGTGAAAAGTAAAGTTCGGTTT 768  
Db 260 CTTGGAGACCTTCATCGGAGGTTCTCCAGCTGCAGGTTCTGGAGGTGACTGTCAAGTT 319  
QY 769 GATGAGGAGCGGATCGAGCGCTGCAAGAGCTATTGCTTACACCTTCAGCGGTCTAATT 828  
Db 320 GTCAAGCGCGCTGTGCTGTGTACAGGAATAAGCGCTTGTTCCTCTCAGCTAAGAGC 379  
QY 829 TTGTAACTGCTTTTCAAAATCGTGCATGCGTCAGTGCCTCCCTCAGCTGGCTTG 888  
Db 380 AAAGTTACAGGATTGAAAAACCTCTGTACAAGGCTCCCGGCTGCAATTGCTCCGG 439  
QY 889 G 889  
Db 440 G 440

## RESULT 5

BM402665

LOCUS

DEFINITION

BM402665 510 bp mRNA linear EST 01-JUL-2002  
SLA006G12.34711 An expressed sequence tag (EST) collection from the  
resurrection plant Selaginella lepidophylla Selaginella  
lepidophylla cDNA clone SLA006G12.5, mRNA sequence.

ACCESSION

BM402665

VERSION

EST.

KEYWORDS

Selaginella lepidophylla

SOURCE

Selaginella lepidophylla

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Lycopodiophyta; Isoetopsida; Selaginellales; Selaginellaceae;  
Selaginella.

REFERENCE

1 (bases 1 to 510)

AUTHORS

Turriaga, G. and Cushman, J.C.

TITLE

An expressed sequence tag (EST) collection from the resurrection  
plant Selaginella lepidophylla

JOURNAL

Unpublished (2002)

COMMENT

Contact: Cushman JC  
Department of Biochemistry  
University of Nevada  
MS200, Reno, NV 89557-0014, USA  
Tel: 775-784-1918  
Fax: 775-784-1650  
Email: icushman@unr.edu  
PCR Primers  
FORWARD: T3 20mer  
BACKWARD: T7 21mer  
Plate: 006 row: G column: 12  
Seq primer: T3 20mer



Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810

Email: estowatson.wustl.edu  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.lnl.gov) for further information.  
Insert Length: 2138 Std Error: 0.00  
Seq primer: -28ml3 rev2 Et from Amersham  
High quality sequence stop: 378.

#### FEATURES

Location/Qualifiers  
1. .488

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:713816"  
/sex="Female"  
/tissue\_type="ovarian tumor"  
/lab\_host="PH10B (ampicillin resistant)"  
/clone\_lib="Soares ovary tumor MBOR"  
/note="Organ: ovary; Vector: pT73D (Pharmacia) with a  
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGCGGTTTCTTTTCTTTT 3'],  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pT73 vector  
(Pharmacia). Library constructed by Bento Soares and  
M. Fatima Bonaldo."

#### ORIGIN

Query Match 3.8%; Score 37.4; DB 9; Length 488;  
Best Local Similarity 49.7%; Pred. No. 17; Indels 0; Gaps 0;  
Matches 95; Conservative 0; Mismatches 96; Indels 0; Gaps 0;  
QY 709 CATGACGAAGTTCGTGGTGGCTTTCTTATGAGGCGCGTGAAGTAAGTTCGGTTT 768  
Db 278 CTTGAGAGCTTCATGCGCAGGTTCACAGTTCGAGTCTGGAGGTGACTGTCAGTT 219  
QY 769 GATGAGGAGCGGCATCGACCGTGCAGAGCTATTCGTTACCTTCAGCGGCTCTAATT 828  
Db 218 GTCCAGCGCGCCCTGCTGCTGTACAGGAAGTAGGCGTGTGCTGGCCAGCGAGGAGC 159  
QY 829 TTGTAACTGCTTTGAAATCGTGAAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGG 888  
Db 158 AGAGTCACCAAGATGGAAGACCTGTGTACAGGCGTCCGCGGTGCTGCTATCCCGG 99  
QY 889 GAAACTGAGGC 899  
Db 98 GCCCCAGGCGC 88

#### RESULT 8

CD875971 560 bp mRNA linear EST 11-JUL-2003  
LOCUS AZ03.10807R011126 AZ03 Triticum aestivum cDNA clone AZ03108B07,  
mRNA sequence.

CD875971  
CD875971.1 GI:32559787

EST.  
Triticum aestivum (bread wheat)  
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.

1 (bases 1 to 560)

Genoplante.  
Genoplante, a major partnership french program in plant genomics  
Unpublished (2003)  
Contact: Genoplante

Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).

#### FEATURES

Location/Qualifiers  
1. .560  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="AZ03108B07"  
/tissue\_type="leaf"  
/clone\_lib="AZ03"

#### ORIGIN

Query Match 3.8%; Score 36.6; DB 14; Length 560;  
Best Local Similarity 62.6%; Pred. No. 31; Indels 0; Gaps 0;  
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 851 GTGAGTCGATGCGTCAGTGGCTCCCTGAGCTGGGAAACTGAGGCGTGGTAGCGG 910  
Db 285 GAGTGTGATGCTGGTGGTTCCTTAGTATGTGACCTGACGCCCTGTTAGCGG 344  
QY 911 AGATCCAGACCATCTGATTCACCTTAACGG 941  
Db 345 AGGATCTCGCGATGAGTCTGCGCCTTGCGCG 375

#### RESULT 9

CD890820 584 bp mRNA linear EST 14-JUL-2003  
LOCUS G118.115108R010926 G118 Triticum aestivum cDNA clone G118115I08,  
mRNA sequence.

CD890820 GI:32560482

EST.  
Triticum aestivum (bread wheat)  
ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Pooideae; Triticeae; Triticum.

1 (bases 1 to 584)

Genoplante.  
Genoplante, a major partnership french program in plant genomics  
Unpublished (2003)  
Contact: Genoplante

Genoplante  
93, rue Henri Rochefort 91025 EVRY CEDEX France  
Tel: 33 1 69 47 54 00

Fax: 33 1 69 47 54 10  
This sequence has been generated in the framework of the french  
plant genomics programme 'Genoplante' (<http://www.genoplante.com>  
and <http://genoplante-info.infobiogen.fr>).

#### FEATURES

Location/Qualifiers  
1. .584  
/organism="Triticum aestivum"  
/mol\_type="mRNA"  
/cultivar="recital"  
/db\_xref="taxon:4565"  
/clone="G118115I08"  
/tissue\_type="grain (118 degrees per day after  
pollination)"  
/clone\_lib="G118"

#### ORIGIN

Query Match 3.8%; Score 36.6; DB 14; Length 584;  
Best Local Similarity 62.6%; Pred. No. 32; Indels 0; Gaps 0;  
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 851 GTGAGTCGATGCGTCAGTGGCTCCCTGAGCTGGGAAACTGAGGCGTGGTAGCGG 910  
Db 318 GAGTGTGATGCTGGTGGTTCCTTAGTATGTGACCTGACGCCCTGTTAGCGG 377

```

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 378 AGGATCTCGCGATGAGCTGCGCCCTTGCGG 408

RESULT 10
LOCUS BG907733 615 bp mRNA linear EST 05-JUN-2001
DEFINITION TaIR1162D06F TaLr1 Triticum aestivum cDNA clone TaIR1162D06 3',
mRNA sequence.
ACCESSION BG907733
VERSION BG907733.1 GI:14315409
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 615)
Miao, G., Caraher, N. and Hanafey, M.K.
DuPont Wheat cDNA Sequence
Unpublished (2002)
Contact: Scott V. Tingey
Crop Genetics
E. I. DuPont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingey@USA.dupont.com
Seq primer: M13.
Location/Qualifiers
source
1..624
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wlin.pk0103.g10"
/tissue type="leaf"
/clone_lib="wlin"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) leaf 7 day old
seedling, light grown (normalized)"

ORIGIN
Query Match 3.8%; Score 36.6; DB 14; Length 624;
Best Local Similarity 62.6%; Pred. No. 33;
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 851 GTGAGTCGATCGTCAGTGGCTCCCTGAGCTGGGAACTGGGAGCGGTGAGCGG 910
| | | | | | | | | | | | | | | | | |
Db 184 GAGTGGTGGTGGCTTGGTGGTTCCTTAGTATGTGACCTTGACGCGCTGAGCGG 125

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 124 AGGATCTCGCGATGAGCTGCGCCCTTGCGG 94

RESULT 12
LOCUS CB902040/c 761 bp mRNA linear EST 02-JUL-2003
DEFINITION trico28xm20 T.reesei mycelial culture, Version 3 april Hypocrea
jecorina cDNA clone trico28xm20, mRNA sequence.
ACCESSION CB902040
VERSION CB902040.1 GI:30116698
KEYWORDS EST.
SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 761)
Foreman, P.K., Brown, D.E., Darkmeyer, L., Dean, R., Diener, S.,
Dunn-Coleman, N.S., Goedegebuur, F., Houfek, T.D., England, G.S.,
Kelley, A.S., Meerman, H.J., Mitchell, T., Mitchinson, C.,
Olivares, H.A., Teunissen, P.J., Yao, J. and Ward, M.
Transcriptional regulation of biomass-degrading enzymes in the
filamentous fungus Trichoderma reesei
J. Biol. Chem. 278 (34), 31988-31997 (2003)
22803314
12788920
Contact: Pamela K. Foreman
Genencor Intl.
925 Page Mill Road, Palo Alto, CA 94304, USA
Tel: (650) 846-7635
Fax: (650) 621-7817
Email: Pforeman@genencor.com
Seq primer: LT-F1 primer.
Location/Qualifiers

FEATURES
source
1..624
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wlin.pk0103.g10"
/tissue type="leaf"
/clone_lib="wlin"
/note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
XhoI; Wheat (Triticum aestivum L.) leaf 7 day old
seedling, light grown (normalized)"

ORIGIN
Query Match 3.8%; Score 36.6; DB 12; Length 615;
Best Local Similarity 62.6%; Pred. No. 33;
Matches 57; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 851 GTGAGTCGATCGTCAGTGGCTCCCTGAGCTGGGAACTGGGAGCGGTGAGCGG 910
| | | | | | | | | | | | | | | | | |
Db 301 GAGTGGTGGTGGCTTGGTGGTTCCTTAGTATGTGACCTTGACGCGCTGAGCGG 360

QY 911 AAGATCCAGACCATCTGATTACCTTAACGG 941
| | | | | | | | | | | | | | | | | |
Db 361 AGGATCTCGCGATGAGCTGCGCCCTTGCGG 391

RESULT 11
LOCUS CA623177/c 624 bp mRNA linear EST 23-NOV-2002
DEFINITION wlin.pk0103.g10 wlin Triticum aestivum cDNA clone wlin.pk0103.g10
5' end, mRNA sequence.
ACCESSION CA623177
VERSION CA623177.1 GI:25201473
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
Location/Qualifiers

```

RESULT 14	AV640730/C	301 bp	linear	EST 15-DEC-2000
LOCUS	AV640730	301 bp	linear	EST 15-DEC-2000
DEFINITION	AV640730 Chlamydomonas reinhardtii 5% CO2 cDNA clone HCL021d11_r 5', mRNA sequence.	301 bp	linear	EST 15-DEC-2000
ACCESSION	AV640730	301 bp	linear	EST 15-DEC-2000
VERSION	AV640730.1	301 bp	linear	EST 15-DEC-2000
KEYWORDS	EST.	301 bp	linear	EST 15-DEC-2000
SOURCE	Chlamydomonas reinhardtii	301 bp	linear	EST 15-DEC-2000
ORGANISM	Chlamydomonas reinhardtii	301 bp	linear	EST 15-DEC-2000
	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadaceae; Chlamydomonas.	301 bp	linear	EST 15-DEC-2000

	Matches	58;	Conservative	0;	Mismatches	36;	Indels	0;	Gaps	07;
Qy	403	GAGGAATCTCTTGTAGTCTTTTC	TGATGAAGCATCGAAGCGGTAAAGCGCGACGGAA	462						
Db	277	GGGGTACTTCTTGCGGTGCTCCTT	GATGACGCGGCCCTTGC	218						
Qy	463	GCTCTCAGGCTGCGATGATCTCT	GTGCGATCTCC	496						
Db	217	GCCTCTCCTCTGACGAAGCCGATG	TAGATCTCC	184						
RESULT 15										
AV638244/c										
LOCUS	AV638244			358 bp	mRNA			linear	EST 15-DEC-2000	
DEFINITION	AV638244 Chlamydomonas reinhardtii 5' CO2 clone HC083g05 r 5', cDNA sequence.							Chlamydomonas reinhardtii		

ACCESSION AV638244  
 VERSION AV638244.1 GI:10781564  
 KEYWORDS EST.  
 SOURCE Chlamydomonas reinhardtii  
 ORGANISM Chlamydomonas reinhardtii  
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 Chlamydomonadaceae; Chlamydomonas.  
 REFERENCE 1 (bases 1 to 358)  
 AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,  
 Nakamura, Y. and Tabata, S.  
 TITLE Generation of expressed sequence tags from low-CO2 and high-CO2  
 adapted cells of Chlamydomonas reinhardtii  
 JOURNAL DNA Res. 7 (5), 305-307 (2000)  
 MEDLINE 20539644  
 PUBMED  
 COMMENT Contact: Erika Asamizu  
 The First Laboratory for Plant Gene Research  
 Kazusa DNA Research Institute  
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.  
 FEATURES  
 Location/Qualifiers  
 1..358  
 /organism="Chlamydomonas reinhardtii"  
 /mol\_type="mRNA"  
 /strain="C9"  
 /db\_xref="taxon:3055"  
 /clone\_lib="HC083G05\_r"  
 /clone\_lib="Chlamydomonas reinhardtii 5% CO2"  
 /note="Vector: pBluescriptII SK-; Site\_1: EcoRI; Site\_2:  
 XhoI; The cDNA library was constructed from cells cultured  
 in a medium with bubbling air containing 5% carbon  
 dioxide"

ORIGIN

Query Match 3.7%; Score 36.4; DB 9; Length 358;  
 Best Local Similarity 61.7%; Pred. No. 28;  
 Matches 58; Conservative 0; Mismatches 36; Indels 0; Gaps 0;  
 QY 403 GAGGAATCTCTGTAGTCTTCTTATGAAGCATCGAAGCGGTTAAAGCGCGAGCGGAA 462  
 |||||  
 Db 278 GGGGTACTCTTGGGGTCTGCTTGTATGACGGGGCCCTTGGGCCCTCGCGGAGCGGAA 219  
 |||||  
 QY 463 GCTCTCCAGGCTGGCATCTCTCTCGATCTCC 496  
 |||||  
 Db 218 GCCCTCTCTTGCAGGAGCGGATGATGATCTCC 185  
 |||||

Search completed: October 2, 2004, 05:12:44  
 Job time : 2828.71 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:25:51 ; Search time 3957.91 Seconds  
(without alignments)  
10644.353 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972  
Sequence: 1 grgaacagcagtgacggcat.....ttgggcgcgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb.ba.\*

2: gb.htg.\*

3: gb.in.\*

4: gb.om.\*

5: gb.ov.\*

6: gb.pat.\*

7: gb.ph.\*

8: gb.pl.\*

9: gb.pr.\*

10: gb.ro.\*

11: gb.sts.\*

12: gb.sy.\*

13: gb.un.\*

14: gb.vi.\*

15: em.ba.\*

16: em.fun.\*

17: em.hum.\*

18: em.in.\*

19: em.mu.\*

20: em.om.\*

21: em.or.\*

22: em.ov.\*

23: em.pat.\*

24: em.ph.\*

25: em.pl.\*

26: em.ro.\*

27: em.sts.\*

28: em.un.\*

29: em.vi.\*

30: em.htg\_hum.\*

31: em.htg\_inv.\*

32: em.htg\_other.\*

33: em.htg\_mus.\*

34: em.htg\_pln.\*

35: em.htg\_rod.\*

36: em.htg\_mam.\*

37: em.htg\_vrt.\*

38: em\_sy.\*

39: em.htgo\_hum.\*

40: em.htgo\_mus.\*

41: em.htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	156.6	16.1	343250	1	AP003594	AP003594 Nostoc sp
2	143	14.7	5039	1	YEH414030	AJ414030 Yersinia
3	135.8	14.0	2892	1	BACHSMR	LO1541 Bacillus su
4	58.8	6.0	3889	1	PROIRM	K02081 P.stuartii
5	45.4	4.7	300511	1	AE016775	AE016775 Pseudomon
6	38.6	4.0	2000	6	AX655393	AX655393 Sequence
7	38.6	4.0	249050	1	AL596165	AL596165 Listeria
8	38.6	4.0	349980	6	AX417038	AX417038 Sequence
9	38.6	4.0	349980	6	AX417041	AX417041 Sequence
10	38.6	4.0	349980	6	AX417042	AX417042 Sequence
11	38.4	4.0	2000	6	AX655393	AX655393 Sequence
12	38.4	4.0	168186	2	AC103403	AC103403 Mus muscu
13	38.4	4.0	177306	2	AC130283	AC130283 Mus muscu
14	38.4	4.0	161360	2	AC132844	AC132844 Mus muscu
15	38	3.9	214511	2	AC129256	AC129256 Rattus no
16	37.6	3.9	4526	1	AF062920	AF062920 Thermus a
17	37.4	3.8	123413	9	AL138834	AL138834 Human DNA
18	37.4	3.8	125020	9	AF429315	AF429315 Homo sapi
19	37.2	3.8	275710	2	AC135276	AC135276 Rattus no
20	36.8	3.8	125020	9	AF429315	AF429315 Homo sapi
21	36.8	3.8	349895	1	BX248359	BX248359 Corynebac
22	36.2	3.7	37106	1	AY228335	AY228335 Xanthomon
23	36.2	3.7	284984	2	AC107121	AC107121 Rattus no
24	36.2	3.7	313488	2	AC128784	AC128784 Rattus no
25	36	3.7	201627	9	AC017068	AC017068 Homo sapi
26	35.8	3.7	12396	1	AE011830	AE011830 Xanthomon
27	35.8	3.7	110000	2	LMFLCHR18_01	Continuation (2 of
28	35.8	3.7	142306	8	AP004343	AP004343 Oryza sat
29	35.8	3.7	184041	9	AC093496	AC093496 Homo sapi
30	35.8	3.7	215657	10	AL592551	AL592551 Mouse DNA
31	35.6	3.7	10211	1	AE012383	AE012383 Xanthomon
32	35.4	3.6	165400	8	OSJN00147	AL662941 Oryza sat
33	35.2	3.6	158449	2	AC115869	AC115869 Mus muscu
34	35	3.6	813	1	AB014991	AB014991 Ralstonia
35	35	3.6	1335	3	AY061932	AY061932 Drosophil
36	35	3.6	1335	3	DME302046	AJ302046 Drosophil
37	35	3.6	57303	2	AC020240	AC020240 Homo sapi
38	35	3.6	65750	9	AC092542	AC092542 Homo sapi
39	35	3.6	142101	2	AC013662	AC013662 Homo sapi
40	35	3.6	171325	9	AC079345	AC079345 Homo sapi
41	35	3.6	176610	3	AC007299	AC007299 Drosophil
42	35	3.6	270775	3	AE003624	AE003624 Drosophil
43	34.8	3.6	1590	6	AR222525	AR222525 Sequence
44	34.8	3.6	1590	6	AR432938	AR432938 Sequence
45	34.8	3.6	1590	6	AX350853	AX350853 Sequence

ALIGNMENTS

RESULT 1  
AP003594  
LOCUS AP003594 343250 bp DNA linear BCT 28-NOV-2001  
DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 14/19.  
ACCESSION AF003594 BA000019  
VERSION AF003594.1 GI:17132824  
KEYWORDS  
SOURCE Nostoc sp. PCC 7120  
ORGANISM Nostoc sp. PCC 7120  
REFERENCE 1  
AUTHORS  
Kaneoko,T., Nakamura,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,  
Watanabe,A., Iriguchi,M., Ishikawa,A., Kawashima,K., Kimura,T.,  
Kishida,Y., Kohara,M., Matsumoto,M., Matsuno,A., Muraki,A.,  
Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,

Yasuda,M. and Tabata,S.  
Complete genomic sequence of the filamentous nitrogen-fixing  
cyanobacterium *Anabaena* sp. strain PCC 7120  
DNA Res. 8 (5), 205-213 (2001)  
21595285  
11759840  
2 (bases 1 to 343250)  
Kaneko,T.  
Direct Submission  
Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research  
Institute, The First Laboratory for Plant Gene Research; Yana  
1532-3, Kisarazu, Chiba 292-0812, Japan  
(E-mail: kaneko@kazusa.or.jp,  
URL: http://www.kazusa.or.jp/cyanobase//,  
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)

FEATURES  
source

1. 343250  
/organism="Nostoc sp. PCC 7120"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:103690"  
/note="synonym: *Anabaena* sp. PCC7120"  
complement (188. .6253)  
/gene="alr3691"  
complement (188. .6253)  
/gene="alr3691"  
/note="ORF ID: alr3691"  
/codon\_start=1  
/transl\_table=11  
/product="serine/threonine kinase with two-component  
sensor domain"  
/protein\_id="BAB75390.1"  
/db\_xref="GI:17132825"

gene

CDS

/translation="MESVKNMTYVDLAIAIPGYFILEBIYHSGKTWVYRAVREADQOP  
WIKLLKREYFPFSELLOPRNOYAIKAKLIPGIVNLXSLPYRNSYAMWEDFGIS  
LRDYAQOCSLITLITITIQADILHHIYQRIIHKDKIPANILINPETHQVKLDF  
STASLLPRETONIISPNIIEGLTAYLSPQGRMNRGIDYSDFSGLVGFELTIGE  
LPFTTDDPELHVCHIAQPKDFTTIPFCQRYANRNGSGESTKLPNEETIPQVGE  
IVMKLMAEDRYQNALGLYDEKLAELGTEGKVEYFTIGSRDCEFRITPDKLY  
GREAEITLALFRVSLGATELMVAGSGIETGATVINEVHKPIVRQGVPIQKFD  
QFORNPISFAQANFLMQLLTSDDTQIOWRSKILEAVGKNGQVIEIPELELI  
IGKQAPAKLSTSLQNFLLFKKPTQVFTSAEHLPIFLDLDQWADSGSLIKLQLL  
MNDTGILLIGYRNEVNPAPHLMLTVEIKGNNGTIIKKDLAPLQIKNTLVAET  
LKCSSETAQSLALLVQTKQGNPFPTTQFLKALHQDNLIQNFYDILGWCDDIAQVNHQ  
ALTDVDFALQKLPRATQEVQLAACIENQPDLETLAIWESSPMKTAACLWKA  
LQEGLIIPQSEYKFEVQOEIIEHAQTEITVYFKLHDRVQQAAYCLIPAEARIAH  
HHIGOLLQKSPAREBHIPEIVNQLNGVITLIQOOSERDELAOLNNAQKAPAT  
AYQASREYADTSMILGTAMCROXQULTIHELAAELALLCGDVEQNNQLLETVMH  
AKCPDRVQVQVKLQALTSNELLEALATGSKFQELGVSPLDPHTPEDVQQAAREI  
NRLIGDRLGELINLPKMTDPEKAIMQISSLIPACYMTGSLLYLLIPIQVKLSLQ  
FNGLSFAGYVSYAFHSTTQNNMALAQQLGQIAYQIASEPEAKNIRAAFTVILGGY  
FYCHTSHLDQTLPIITQEGCIALETGNLEFFGVYQNFQNMAYWGESLSPASOMCT  
YHQQLLDENLVTSAMHCFVYVESALILIGKSADEISLRQDAYAEKIIVTEALASNDLWR  
LFOFVLYRLVLNFIILADPRAKDAEKARQYLTGCMGSAEPIEYVDSLTALAEHLQ  
PAGBDSQORVRENQALIDKADHAPMNYLHKWQVAAETHRLLEQKTEAMEFTEMAI  
KGAKENLYDEALANLAAKPIYLDWGKEKVAAYVMQGEIYCYARKEAAKINDLEK  
YPQLQPIILQREKFNFEETIYVGTSSSTLSSTGTSITSDTDFASIKLAAQVIS  
TSLEDEILTNLTIELNSGAKKSALLIPQDIWQIKAMTILSNPNSSESQTILE  
SQPLETCDIPKNIYIYVKNYQQTVIDNLQTDIPGLIEYMLQHOPOSFCTPMNQ  
GHVGLILVLENLITRGVTSRLEVIIRLSQAQAVSLKARLYOESOTKAQOLKSLX  
QOKLIFNVNQMRQSLDNLNAFCVVTQNIIRILLVDVRGIIYQFHLDDVNYEIGFVAED  
VSPAPPSALAVQVQDHCGENYANDYKQRCIATDVQSSILDCRQILQAQHVRAE  
LVPIMQBEELWGLCICHCQDRPQWELEQFAQVQAGMGIKQTDLLIQTOKQA  
TQLEHTLQLOQTQQLVQNEKMSALGNLAVAGHMMNPLGFTAVSLQHTQPTFADI  
MEHLKYQESLPHQDKILHAAQIDLDYILEDLPLKTEAMVACDRKNTSTSLRTF  
SRADKYKIPFNHKKIDSTLILKHLKADKRAPIEIVTEYGNLPHIECFPGOLNQ  
VFMNLNADIALEDANNGLSVAETANNRILIRTTQVNNHVRISITDNGIGMGEEL  
KQHLFEHLFTTKGVEKGUGLGLAIARQIVKXHGGAIEVNSQLQGGTGFVLIIFPITAE  
SANKY"

gene

CDS

hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75391.1"  
/db\_xref="GI:17132826"  
/translation="WNFPIQGVVALALLTQGLKVLGAIALWIVAKLHIEFALKGVRR  
GLRTQNIETPLIISYLLINIVSVTRIVLIVAILGFFGIETTSFAALLAAGVAIGAWR  
GLLANFAAGAFLLIIFRPKVGDFITAAAGITGVSBIGLFTTIDITPDNVLTIVANNKI  
PADNIONFSANPFRVDDLLAQLHHDVHNOAIALLKARIISOIPNIENPAPDEVILTF  
NLAGPYLAVRPPYCNNDHWQVYFDINKAIRETFGEAGYPIPEQRVGFSPENGTSV  
IPSSALSS"  
8079. .8939  
/gene="alr3693"  
8079. .8939  
/gene="alr3693"  
/note="ORF ID: alr3693"  
hypothetical protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75392.1"  
/db\_xref="GI:17132827"  
/translation="MNVAEISQVALALLTQGLRVLGAIALWIIAQLLINFGKLKLR  
APRSKHVPTLINIVNVDVILKIVLIVAILGFFGIETTSFAALLAAGVAIGAWR  
GLLANFAAGAFLLIIFRPFAVGDTIKAAAGTVKVEIGLFTTIDITPDNVLTIVANNKI  
PADNIONFSANPFRVDDLEAQLHHDVHNOAIALLKARIISOIPNIENPAPDEVILTF  
NLLGTVLAVRPPYCHHEHYVQVFDINKAIRETFGEAGYPIPEQRVAFVTCSPDGTSA  
VVORTSLGRG"  
9664. .9831  
/gene="asr3694"  
9664. .9831  
/gene="asr3694"  
/note="ORF ID: asr3694"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75393.1"  
/db\_xref="GI:17132828"  
/translation="MAVDLDKRLVLILQSFHLRSWERNGLWEVSRRIKRBHTHVSM  
VAPETELVPG"  
complement (9841. .11007)  
/gene="all3695"  
complement (9841. .11007)  
/gene="all3695"  
/note="ORF ID: all3695"  
/codon\_start=1  
/transl\_table=11  
/product="N-acyl-D-glucosamine 2-epimerase"  
/protein\_id="BAB75394.1"  
/db\_xref="GI:17132829"  
/translation="MGKNLOALQLYKNALLNDVLPFWNEHSLDTEGGYFTCLDRQK  
VYDTRFTWLNQVQWTFSMCNOLKRENNLKIASGAKFLAQRGRAGNWYFALT  
RGGPRLVQYNIIFSDCFRAAFSQVALASGEWAKDVAMQYNNVLRKDPKGYTK  
TYGTRPMKALAVPMILANLMEWELLPQSTLENLVSATVOEVMDSFLDKRGLMWE  
NVADPMKIDCFEGRILNPHGIEIAMFIMDIARRKNDKSTINQAVDVLNLLNFAE  
KEYGGLYFDAAAGHPQOLEWDQKLMMVHLESLVALMGYRLTGRVYCWYQWHD  
YSWGHYADPEYGEFGVYLNRRGEVLLALKGGKWKGCFFHVPRAALYLCWQCFALS"  
complement (11134. .12273)  
/gene="all3696"  
complement (11134. .12273)  
/gene="all3696"  
/note="ORF ID: all3696"  
unknown protein"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="BAB75395.1"  
/db\_xref="GI:17132830"  
/translation="MLFGAIALDNGRLLVAENGTLOQVLIYKIQDPVQVGTFGHTG  
GIYQGVFGEVDLQFYGLTGVGDAQGNLYINNNFNNSGTDLRKFSQGLWELLG  
LSLVENADAPKTDGREVFTYQEQFLIDSSPKGQQLIYKAYTLNGFKYQPSRLHIS  
PDITFVRINDPERFLDLPYSNALEIYRNPNSKQGLITAI PAGMFGVGNIRKSAISG  
TWPEQPTGKMIWDRNGDCAFDEKEDSESDSSYGMWVSGDVWTLTGDRGI  
RHYPLQGLDNGNPIYNTSQEQETELIFDRLRIKYPESDNTNLYSGTAKNPPFA  
EDQAPGSEIACFDNWSKDNRLRIVVPKDTIRKREMITASTH"

gene	complement (12278..13114) /gene="all3697"
CDS	complement (12278..13114) /gene="all3697" /note="ORF ID: all3697 unknown protein" /codon_start=1 /transl_table=11 /protein_id="BAB75396.1" /db_xref="GI:17132831" /translation="MKHEILDFKRLQSRKQSNKIFKYLRYLLKSWIFLIILAIIS FLAIFITADLILSEVAENSDKOLMDIPADSKSIONVIEGLYVAPNGTVVTSNHS DRAGAPASYIKGNVIGVLSLDGSRGKAVTANSKYIYIAMSOGFVGKIDKQYPP EGTWTYVRNYNLSPKPFANGRWKSNMLVNTKSEVTGLATKDKNLYVSDAANK IRYVNTETWKELENFSLRPEDITIDKQGNLWIIQKODANNFAKILIRYSSGKQLPQK NN"
gene	14028..15197 /gene="hepB"
CDS	14028..15197 /gene="hepB" /note="ORF ID: alr3698 hepB gene product" /codon_start=1 /transl_table=11 /product="heterocyst envelope polysaccharide synthesis protein" /protein_id="BAB75397.1" /db_xref="GI:17132832" /translations="MEYKKEKLNLSVASILITLGSWFPPTNPGGLERYIYELTYQLSAN ODRVELGCVLPNQHPLKLTNLASPDSKIWOPFWSIRNFKOTRIGKPDAINLHF ALYSFPLDILPOGIPITFNPHGPWASESKOELVKNKISIFLAKRLIEQOTYVNHCDRF IVLSKAFGNILHQYQIIPWKHIHIIIPGVNIDIFQPNLSRQQRQLNWPESRPLPT SRRLVRVGVKLLQALIKPLDILWLAIGRHLLQTLLEKQALGLENVKELG FLDEQIPIAYQANLTMPSQSFEGFLAITESLACGTPLVICTPIGMPPEILTPFPSP QLITASPEATAIEAKIAILLEQIPKPSRECRQYAVTNFEDWOKIAQQVRQVILA" 15259..16407 /gene="alr3699"
gene	15259..16407 /gene="alr3699"
CDS	15259..16407 /gene="alr3699" /note="ORF ID: alr3699 similar to glycosyltransferase"
Query Match	16.1%; Score 156.6; DB 1; Length 343250;
Best Local Similarity	49.6%; Pred. No. 4.6e-34;
Matches	461; Conservative 0; Mismatches 459; Indels 9; Gaps 2;
QY	38 GCATCGATACGCGGCGCTGCTAAAGCGCTTTGGGTTTGACGCGCAACGATATAACG 97
Db	18052 GAATTGAGGCGGTTGGCAATTTCTTAAGATATTTCTGCTCCCAAGACGACAGAACG 18111
QY	98 TCCGTAGCGCTGTGACATTTGTCGCGCTTGCGGTTTGAAGCGGGAGATCGCTGGTTG 157
Db	18112 AGCGTTCAGACATTTCTTACTTGCATTGGCAGATATTCGTCGCGAAATTCCTTGGACTC 18171
QY	158 ACTGACACATCCAGCCCTTGGCGTTTCAAGACATCATGACTGTCGCGGAGCATTCGG 217
Db	18172 AAGCTATACATCCAGACGCGAATTTACAGAAATGATGATGGTTGGTTCTGTATCTACTATG 18231
QY	218 CCAAGCGCTAGCCACCGGAGTCCAGAGATTTCCGTTAAGAGACGCTTCGGCAGTGGG 277
Db	18232 GTAACAATAGCCACCGAATACACCGGAACNGTCAGACGCGCAGCATGACCAAGTTTG 18291
QY	278 TTGATAATGGCTTCCGCGTACTTAATGCGGACAAATTTAAACATCGCGACGAATCGCGAGC 337
Db	18292 TCCAGATGGGATAGTTGTTGAGATCCAGATCAACCCAGATCGACCGGATTAACATGCCA 18351
QY	338 TCAACGAGTACTGCTTGTCAGAGCATTTACAGCGCTAGGCGCATATGAAACGGAG 397
Db	18352 AATGGTGTATCAACTTCATCAACAATTTGTGACACTGCTTAAATACACGCTCCGAGC 18411
QY	398 GCTTCGAGGAATCTCTGTAGTCTTTCTTGATGAAGCATCGAAGCGGTTAAAGCGCGAG 457
Db	18412 AGTGGAGAGACTCGTCGTAATATGTTATTTCAGTTAAAACTTATTGCGAGCAGAA 18471
QY	458 CGGAAGCTCTCCAGGCTGCGATGATCTCTGTGATCTCCCTGCTGCGGAGGAATTTCTGC 517
Db	18472 ATCGAATAATCC-----CGATGATTCGCTAAGTTTGGCTAACGGTCAAGCAATTCAC 18525
QY	518 TCTCGCCTCCGCGGAGAAATCCATTGCTGAAAGATGCTGGAAGATTTGTGCGCGCAT 577
Db	18526 TATCATCAGTGCACAAAATATATTGATTAAGAAATCTGGAATACTTTGTGCCAGAT 18585
QY	578 TTGCACCTCGCTCGAGGTCCTTACTCTCGGGGATCTCGTGAAGACATTCCTCTATTCCG 637
Db	18586 TTACACCTGAAGTTTGTGTTTGTAGGTGATCTCGAATAAAGTTTATTGTTAATG 18645
QY	638 AACGAGAGATCTTTGAAGAGGTGCTCGCGCTGACTTTTCGACCCCAATGCTGGAATGCCGG 697
Db	18646 AGACTCAGAGTTTCAGAAATTT---GGGATCGATTAGATCTCTACGGTAAAGATGCCAG 18702
QY	698 ACCTTATTTCCATGACGAAGTTCTGGGGTGGCTTTTCTTATGAGGCGCGTGAAAGATA 757
Db	18703 ACATTGTGGTTTACTACGAACGCGAAGAGTGGCTTGTCTTTGATAGACGCGTGACTAGTC 18762
QY	758 AAGGTCGGTTGATGAGGCGGCGCATCGAGCCTGCAAGAGCTATTTCGTTACACCTTCAG 817
Db	18763 ATGGTCGGTCAATCTAAACGTCGCAACGAATTAAGCGGCTCTTTCATCTAGCCGTC 18822
QY	818 CGGGTCTAAATTTTGTAAACTGCTTTGAAATCGTGAGTCGATCGGTCAGTGGCTCCCTG 877
Db	18823 AGGGTTAGTTTTCGTTACTGCTTCCCGAGTCGAAGAGAAATGACTCGCTATCTTCTG 18882
QY	878 AGCTGGCTTGGGAACTGAGCGGTGGGTAGCGAAGATCCAGACCATCTGATTCACCTTA 937
Db	18883 AAATCTCTTGGGACAGAAAGTTTGGGTAGCAGCTCAACCTGATCAATGATTCATTCA 18942
QY	938 ACGGCTTACATTTCTTGGCGGTACGAA 966
Db	18943 ATGGAAGAAAGATTTCTCGGACCCCTATGAA 18971
RESULT 2	
YEN414030	5039 bp DNA linear BCT 27-MAR-2003
LOCUS	Yersinia enterocolitica Yei gene and ORF1 and insertion sequence
DEFINITION	IS1222 ORFA (partial) and ORFB.
ACCESSION	AJ414030.1 GI:17977862
VERSION	AJ414030.1
KEYWORDS	methylesterase-endonuclease; ORF1; ORFA; ORFB; transposase; Yei1 gene.
SOURCE	Yersinia enterocolitica
ORGANISM	Yersinia enterocolitica Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Yersinia.
REFERENCE	1 Antonenko, V., Pawlow, V., Heesemann, J. and Rakin, A. Characterization of a novel unique restriction-modification system from Yersinia enterocolitica O:8 1B FEMS Microbiol. Lett. 219 (2), 249-252 (2003)
JOURNAL	2508023
MEDLINE	12620828
PUBMED	12620828
REFERENCE	2 (bases 1 to 5039)
AUTHORS	Rakin, A.V.
TITLE	Direct Submission
JOURNAL	Submitted (21-SEP-2001) Rakin A.V., BAF, Max von Petttenkofer-Institute, LMU, Pettenkofer Str. 9a, Munich, 80336, GERMANY
FEATURES	Location/Qualifiers 1..5039 /organism="Yersinia enterocolitica" /mol_type="genomic DNA" /strain="8081" /db_xref="taxon:630" /country="USA" 1..432 /transposon="IS1222"
repeat_region	

CDS	complement (<1..71) /function="essential for IS1222 transposition" /note="ORF1" /codon_start=1 /transl_table=11 /product="transposase" /protein_id="CAC95148.1" /db_xref="GI:17977863" /db_xref="SPTREMBL:Q8VLS1" /translation="MLMCDATGLSQRACELTGLSLST" complement (98..361) /function="essential for IS1222 transposition" /note="ORF2" /codon_start=1 /transl_table=11 /product="transposase" /protein_id="CAC95149.1" /db_xref="GI:17977864" /db_xref="SPTREMBL:Q46611" /translation="MKKRFSDQIIISILREAEAGVSARELCRKHAISDATFTWRKKY GMEVPEVRLKSLSEENARLKLLAEMLDKEALQVALGRKY" 389..432	repeat_unit	93	TAACGTCGCTAGCGCTGTGACATTCGCTCGCGCTTGCCTGCGGTTTGAAGCCGGGAGATCGCTG	152
			2707	AAATGAGCGTTTCAGCTCTTACATTTCTTGACATAGTCAACCTAAGACCAAGAGTTTCATG	2766
			153	GGTTGACTCGACCACTCCACGCTTGGCGTTTCAAGAGATCATGACTGGTCCGCGGAGCA	212
			2767	GCAGGAATTAGAAAAACCTTTGGTTGGTTACTCCAAATTATGGATTGGTTGATAGATAT	2826
CDS			213	TTGGGCAAGCCGTAACGCAACCGAAGTCGAGAGATTTCCGTAAGAAGACGCTTCGGCA	272
			2827	ATATGCAAAAGATATGCGCTAATACGAGAGAAACATTTAGACGCGCAACATTCACACCA	2886
			273	GTGGGTTGATATGGCTTCGCGTACTTAATGCGGACAAATTTAAACATGCCACGAACTC	332
			2887	ATTATTTGATGGTGGACTCGTCTTTACAAACCCAGATAAACCCCAATCGTCAGTAAATAG	2946
			333	GCAGCTCAACGAGTACTGCTGTCTGACGAAGCATTTACAGCGCTAAGGCGCATATGGAAC	392
			2947	TCCTAAGCTTGTATCAATTTGCACCAAGATTTATTTGATGATATGATACCTATGGGAC	3006
			393	GGAGGCTTCGAGGAATCTCTTGTAGTCTTTCTTGATGAAGCATCGAAGCGGTTAAAGC	452
			3007	TCCTCTCTGGAATAAAGCATTTAGGTGAGTGGCTAATGTCAGCGAGAGACATTTGTCGAAC	3066
			453	GGAGCGGAAGCTCTCCAGGCTGCGATGATCTCTGTCGATCTCCCTGGTGGCGAGGAAT	512
			3067	ATATGCCATGAAGCGTGAATCATATGATTTCTCTGACCATCGATAATTTGTTACTGAGAT	3126
			513	TCGTCTCTCGCTCGCGGCGAGATTCATTTGCTGAAGAAGATGGTCAAGAGTTTCTGCG	572
			3127	CCACTTAAGTCCCGGTGACCATGATCAATTAATACATGATATCGTTACTGAAATTTGGGC	3186
			573	GGGATTTGCACTCGCTGCGACGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	632
			3187	TCGATTTGCGCGAGGCTCTCAAGTTATATATTTAGCGACACTGCTGCTGCTGCTGCTGCT	3246
			633	ATTGCAACGAGAGATCTTTGAGAGGTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	692
			3247	TTTCAGAAAGACGATTTAGCAGAT---CTTGAGTTACTGTTAATCTGTAAGGGGAACT	3303
			693	CGCGACCTTATTTCTCCATGACGAAGTTCTGCTGGTGGCTTTTCTTATGAGCGCGTGA	752
			3304	TCCTGACGTTGTTCTGATTTGCGCCGACGAGATTTGTTAATCTTATAGATCAGTAC	3363
			753	AAGTAAAGTCGTTTGTATGAGAGCGGATCGACCTGCAAGAGCTATTCTGTTACACC	812
			3364	TAGTCATGCGCTGTCGATGGAAGACATAGCGAACTTGCAAACTTTTCAAGGATGC	3423
			813	TTGACGCGGTCTAATTTTGTAACTGCTTGAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	872
			3424	TCGCGCAGGATTTGTTATGATCTGCTTTCCCGGATAAAAAACAATGAGTAAATCTT	3483
			873	CCCTGAGCTGGTTGGGAACTGAGCGGTGAGCGGATGCGGAGATTCAGACCATCTGATCA	932
			3484	CTCAGAAATAGCTGGGAAACAGAGTTTGGATAGTCTGAGCTCGACCCATATGATCA	3543
			933	CTTAAACGGTCTAGATTTCTTGGGCC	959
			3544	CCTAAATGGCGACAGGTTCTTAGGTCC	3570
RESULT 3					
BACHSMR					
LOCUS	BACHSMR	2892 bp	DNA	linear	BCT 11-FEB-2002
DEFINITION	Bacillus subtilis modification methyltransferase (hsmBI) and restriction endonuclease (hmrBI) genes, complete cds.				
ACCESSION	L01541	S52585			
VERSION	L01541.1	GI:143053			
KEYWORDS	hsmBI gene; hmrBI gene; modification methyltransferase; restriction endonuclease.				
SOURCE	Bacillus subtilis				
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				

REFERENCE 1 (bases 1 to 2892)  
 AUTHORS Xu, G.L., Kapfer, W., Walter, J. and Trautner, T.A.  
 TITLE BuBI--an isospecific restriction and modification system of PstI:  
 characterization of the BsuBI genes and enzymes  
 JOURNAL Nucleic Acids Res. 20 (24), 6517-6523 (1992)  
 MEDLINE 93126092  
 PUBMED 1480472  
 COMMENT On Feb 8, 2002 this sequence version replaced gi:263379.  
 FEATURES  
 source Location/Qualifiers  
 1..2892  
 /organism="Bacillus subtilis"  
 /mol\_type="genomic DNA"  
 /strain="ISB8"  
 /db\_xref="taxon:1423"  
 185..190  
 /gene="hsmBI"  
 198..21733  
 /gene="hsmBI"  
 214..220  
 /gene="hsmBI"  
 228..1733  
 /gene="hsmBI"  
 228..1733  
 /gene="hsmBI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="modification methyltransferase"  
 /protein\_id="AA18169.1"  
 /db\_xref="GI:143054"  
 /translation="MTQILETVDKSRITVNLKNSKELQGFPPSSISIFMACLFSE  
 DKLNNKVLDDAGVIGSLTAFILARLISENIGRADLHLEIDEMLEPYLSTLALPKD  
 YEINQIILDDFWEAYASLLDDESLANDKORFTHAILNPVKYIKSKSRKLLR  
 KAGETVNLVSAFVALVDLMSDGEIVFVLPSPGCPYRPHRQHLNKTSLKXNH  
 LPSRDKAFKDDEVLQENLVSKLEKIVQSDKISIDSDS SVIRYRYPFEKIVOP  
 LNFQRKTHINTMETLIEKHNPVCSLEELNEIVSGPVDFRVKENVLEMPGEGTV  
 PLFYPHFVTSLEYEIPOMKKPNAIIRNEKVKWLYPNHGVVYVVKRFSSEKERRIVA  
 GVLTPEVDVDFEGLNVLHYNKSGISKVAYGLYVNLSTPVDKYPRIENGHTQ  
 VNATDLRTMKPPSRDLILSLGWVNIENVGVEIDSLEELLNDRGNA"  
 1697..1701  
 /gene="hsmBI"  
 1714..2664  
 /gene="hsmBI"  
 1714..2664  
 /gene="hsmBI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="restriction endonuclease"  
 /protein\_id="AA18170.1"  
 /db\_xref="GI:143055"  
 /translation="MTEGHSNVKEAIKILKELGPKQONERSALCLLSLMTQDK  
 TWSEAPSLIGITPMMEFCRINRYKEVAPNSRETFRTMHQFVDAGIALYNPKPFR  
 PVSFPAVQIEAETLEIKCYNTEWSELLARYLNSRQTLVRYAKEROQNKIPVQI  
 AEGKHYITPGEHSELIKALIEFAPRYVPGRLIYAGDTGEKMGVDFEELLRLGVV  
 IDSHGKMPDVYVIFPEKKWLLILLESVTSHGPDVHKHEELAKLFNGSTAGIVVTAPP  
 NRSMLARYLNINSWETEVWVADAPSHLIHFNVRFLGPYE"  
 19 ATCGACGGAACGGTAGCGACATCGATCTGCGCGGCGCTGCTAAAGCGTTTGGGTTT 78  
 1717 ACCGAGGATGCATAGTAACTGCTCAAGAGACATATAAATCTTAAAGATTAGTTTG 1776  
 79 GACCGCAACGATATAACGTCGCTAGCGCTGTGACATGCTCGCGCTTGGCGTTTGAAG 138  
 1777 CCTAAGGGGAGCAAAATGAACGTTACGCTCTATGTTTATCTTAAATGAATAACT 1836  
 139 CCGGAGATCGCTGGGTTGACTGACCACTCCAGCCCTTGGCTTGAAGATCATGGAC 198  
 1837 CAAGATAAGACATGGTGCAGAGCTGAAAGTCTCTTATTGGTATAACCAATGATGAA 1896

## ORIGIN

Query Match 14.0%; Score 135.8; DB 1; Length 2892;  
 Best Local Similarity 47.2%; Pred. No. 3.9e-28;  
 Matches 447; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

Qy 199 TGGTCCGCGAGCATTGGCCCAAGCCGTACGCCACCGGAACTCGAGAGATTTCGGTAAG 258  
 Db 1997 TTTTGGCGAATTAAATATGGAAGAAGATATGCAACAATAGTCGAGAAGCTTTTAGAAGG 1956  
 Qy 259 AAGACGGCTTCGGCAGTGGGTTGATAATAGGCTTCGCCGTACTTAATTCGGGACAAATTTAAAC 318  
 Db 1957 TTCACTATGCATCAGTTTGTAGATGCTGGAATTCGTTGTATATACCCCTGATAAACCGACA 2016  
 Qy 319 ATCCCCAGAACTCGAGCTCAACGAGTACTGTGCTGACGAAAGCATTTACAGCGGTA 378  
 Db 2017 AGACCTGTTAATAGTCCGAAGCGCAGTATATCAAAATAGAAAGCGGAACTTTAGAACTTATT 2076  
 Qy 379 AGGSCATATGAACGGAAGGCTTCGAGGAATCTCTGTAGTCTTCTTGTATGAAGCATCG 438  
 Db 2077 AAGTGCTATATACAGAAGAAATGAGTGAATTAAGCCCGATCTATCTATTAATAGCAA 2136  
 Qy 439 AAGCGGTTTAAAGCGCGAGCGGAAGCTCTCCAGGCTCGATGATCTCTGTGATCTCCCT 498  
 Db 2137 ACATTGCTAGAAAGGTATGCTAAAGAGAGACAGCAGATAAATACCTGTACAATAGCT 2196  
 Qy 499 GGTGGCGAGGAATTTCTGCTCTGCCCTGCCGCGGAGAAATCCATTTGCTGAAGAAAGTGGT 558  
 Db 2197 GAAGGTTAAAGAGATTTATATAACTCCAGAGAGACACAGCGAGTTAATTAAGGCAATTATA 2256  
 Qy 559 GAAGAGTTTGTGCGCGATTTGCACCTCGCTCGAGCGGTCTCTACCTCGGGGATACCTCGT 618  
 Db 2257 GAAGAATTTGCACCCCTTATGATACAGGAGGAGCATTAATTTATGCTGGTGATCTGGT 2316  
 Qy 619 GGAAGCATTTCCCTATTTCGAACGAGAGATCTTTGAAGAGGTGCTCGGCGCTGACCTTCGAC 678  
 Db 2317 GAAAGATGGTATTATTTTGTAGGAATTAATTAAGACAGTTA---GGGTAGTTATAGAC 2373  
 Qy 679 CCCATGGTCGAATCCGCGACCTTATTTCTCCATGACGAAGTTCTGGTGGTGGCTTTTCTT 738  
 Db 2374 TCGCATGGAAAAATGCCTGATGTAGTAATTAATTTTCCAGAGAAAAAGTGGCTCCTACTC 2433  
 Qy 739 ATGAGAGCCGTGAAAGTAAGTCCGTTTGTAGAGAGCGGATCCAGCGCTCGAAGAG 798  
 Db 2434 ATTGAATCGGTAACTAGTCACGGTCCAGTAGACCATTAAGCGTCAAGAAGTAATGACCAA 2493  
 Qy 799 CTATTCGTTACACCTTCAGCGGCTTAATTTTGTAAACTGCTTTGAAATCGTGAGTCG 858  
 Db 2494 TTAATTAATGTTCAACTGCGGGAATAGTTATGTGACTCAITCCCTAATCGTTCAATTA 2553  
 Qy 859 ATGGTCAGTGGCTCCCTGAGCTGGCTGGGAACCTGAGCGTGGGTAGCGGAAGATCCA 918  
 Db 2554 ATGGCAAGATACCTAAATAACATTTCTTGGGAGACGGAAGTGTGGTAGCTGATGCACG 2613  
 Qy 919 GACCATCTGATTACCTTAACGGGTCTAGATTTCTTGGCGCGTACGA 965  
 Db 2614 TCACATCAATTCATTTTAATGGGTTTCGTTTCTTAGGCCCATACGA 2660

PROISM 3889 bp DNA linear BCT 16-FEB-1994  
 DEFINITION P.stuartii PstI restriction and modification genes, complete.  
 ACCESSION K02081  
 VERSION K02081.1 GI:150922  
 KEYWORDS endonuclease; methylase; modification enzyme; restriction endonuclease.  
 SOURCE Providencia stuartii  
 ORGANISM Providencia stuartii  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Providencia.  
 1 (bases 1 to 3889)  
 AUTHORS Walder, R. Y., Walder, J. A. and Donelson, J. E.  
 TITLE The organization and complete nucleotide sequence of the PstI restriction-modification system  
 JOURNAL J. Biol. Chem. 259 (12), 8015-8026 (1984)  
 MEDLINE 84239756  
 PUBMED 6330092  
 COMMENT Original source text: P.stuartii DNA, clone pPst201.

The two genes for the PstI restriction-modification system are encoded on opposite strands and are transcribed from separate promoters. The restriction endonuclease contains 326 amino acids. The modification enzyme contains 507 amino acids. The close proximity of the transcription initiation points of the two genes (70 bp) suggests that the promoters overlap. The -10 regions for the two genes are found at bp 2307-2302 (MME) and 2352-2357 (PstI). Bases 2267-2340 are protected from DNaseI digestion. The PstI restriction-modification system is a type II system. The two enzymes both recognize the hexanucleotide sequence 'ctgcag'. The modification enzyme protects this sequence from cleavage by PstI by methylation of its adenosine residue. PstI cleaves the DNA between the adenosine and guanine residues, leaving 'sticky ends' of 4 bp.

FEATURES  
Location/Qualifiers  
1..3889  
/organism="Providencia stuartii"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:588"  
complement (<742..2292)  
/product="MME mRNA"  
complement (742..2265)  
/note="methylase modification enzyme (MME)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA25672.1"  
/db\_xref="GI:455323"  
/translation="MTEAATOLPISILNIVDSIRAA NSTLDLTLRSKLGQFMSSSAV  
SELMANLESYVGEHILDAGVGSILTAFAVONATLNGAKSISCTCYEISRVYVNL  
IQVLDLCKRMEFVFNWQKILIESDFOASVEQLLIENYSPKYNKAILNPFYKIAA  
KGRALDQKIVGSIASVAFVLAIAIKOLKSGSELVAITPSPNGPFFDFRQML  
DECSLNKHIVNSRFSKAFKADNLQENIYHLTKGTQRKVTVVSTCANDINTIF  
VPFDEIVKSNPDLFIHIVTNEQRELAKAGGLPCSLDGIQVSTGVNDFFRTRE  
NLSMEYISNVLPIPOHLOKRSIVWPTTKAKPNALIVNEATNLPNGIYVLTTR  
LTAKREKRIVASIVPLDANVTGDFONKYNFANGKPLDLSIAGLWFLNLTLL  
DKYFQMGHTVONADLRALRYPTREQLIEDIANQVDFGEFEQTKIDELINQSLQIM"  
2363..23376  
/product="PstI mRNA"  
2396..23376  
/note="PstI restriction endonuclease (PstI)"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAA25673.1"  
/db\_xref="GI:150924"  
/translation="MKELKKEAKEIKALGLPPOQYNDRSGVLLALANIKPESWK  
EAKPLLPVIMEIRTYGDKYKPNSETIRRTLHOFQEARIVDRNRDLPSPRATN  
SKDNYSLNVLIDILHVPNGNWKELIQOFLTHVPSLQELYERALARDRIPKILIDG  
TQISLSPGHNLHADIHVEFCRFVGMGKILYIGDTASSRNEGKLMVLDSYLKK  
LGVFPMHDKLPDVVYDEKRWLFLEIAVTHSGPISPKRWLELEALSSCTVGVYIV  
TAPPTTFEKRANAIAMETEVMIADNPDHMFNGDRFLGPHDKKPELS"  
1 bp upstream of HindIII site.

ORIGIN 1 bp upstream of HindIII site.  
Query Match 6.0%; Score 58.8; DB 1; Length 3889;  
Best Local Similarity 50.3%; Pred. No. 1.4e-05;  
Matches 144; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 680 CCCATGGTCCGACGACCTTATCTCCATGACGAACTGCTGGGTGGCTTTCCCTTA 739  
Db 3069 CACACGACAAITGCCAGATGTTGTTGTACATGAAAGCGTAAATGTTATTTTAA 3128  
QY 740 TGGAGCCGTGAAAAGTAAAGTTCGTTTGTAGAGAGCGGCATCCACCGTCAAGAGC 799  
Db 3129 TAGAAGCGGTTACATCCCATGCGCCCAATTAGTCTCTAAACGTTGGCTTGAAGCTG 3188  
QY 800 TATTCGTTACCTTCAGCGGGTCTAATTTTCTAACTGCTTTGAAATCGTGAGTCGA 859  
Db 3189 CCTTGTCACTCACTGACGTCGGTAAAGTGTATGTTACTGCTTCCCAACACGTACAGT 3248  
QY 860 TCGCTCAGTGGTCCCTGAGCTGGCTGGGAACTGAGCGGTGGGTAGCGGAAGATCCAG 919  
Db 3249 TCAGAAAAAATCGGCAAAATTGTCATGCGAAACCGAGGTTGATTGCTGTATAATCCAG 3308  
QY 920 ACATCTGATTACCTTAACGGTCTAGATTTCTTGGCGGTACGA 965

Db 3309 ATCATATGGTCCATTTCATTAATGGCGACCGTTTCTTAGGCCACATGA 3354

RESULT 5  
AE016775/c  
LOCUS  
DEFINITION  
Pseudomonas putida KT2440 section 2 of 21 of the complete genome.  
ACCESSION  
AE016775 AE015451  
VERSION  
AE016775.1 GI:26557018  
KEYWORDS  
Pseudomonas putida KT2440  
SOURCE  
Pseudomonas putida KT2440  
ORGANISM  
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
Pseudomonadaceae; Pseudomonas.  
REFERENCE  
1 (bases 1 to 300511)  
AUTHORS  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Uterback,T.,  
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,  
Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,  
Duesterhoft,A., Tummler,B. and Fraser,C.  
Complete genome sequence and comparative analysis of the  
metabolically versatile Pseudomonas putida KT2440  
Environ. Microbiol. 4 (12), 799-808 (2002)  
2 (bases 1 to 300511)  
AUTHORS  
Nelson,K., Paulsen,I., Weinel,C., Dodson,R., Hilbert,H., Fouts,D.,  
Gill,S., Pop,M., Martins Dos Santos,V., Holmes,M., Brinkac,L.,  
Beanan,M., DeBoy,R., Daugherty,S., Kolonay,J., Madupu,R.,  
Nelson,W., White,O., Peterson,J., Khouri,H., Hance,I., Lee,P.,  
Holtzapple,E., Scanlan,D., Tran,K., Moazzaz,A., Uterback,T.,  
Rizzo,M., Lee,K., Kosack,D., Moestl,D., Wedler,H., Lauber,J.,  
Hoheisel,J., Straetz,M., Heim,S., Kiewitz,C., Eisen,J., Timmis,K.,  
Duesterhoft,A., Tummler,B. and Fraser,C.  
Direct Submission  
Submitted (05-NOV-2002) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
Location/Qualifiers  
1..300511  
/organism="Pseudomonas putida KT2440"  
/mol\_type="genomic DNA"  
/strain="KT2440"  
/db\_xref="taxon:160488"  
complement (80..985)  
/gene="rimK"  
/locus\_tag="PP0248"  
complement (80..985)  
/gene="rimK"  
/locus\_tag="PP0248"  
/note="similar to GB:X57130, and PID:296288; identified by  
sequence similarity; putative"  
/codon\_start=1  
/transl\_table=11  
/product="ribosomal protein S6 modification protein"  
/protein\_id="AAN65880.1"  
/db\_xref="GI:24981607"  
/translation="MKIIVLSRNPRLYSTRRLVEAGTQRGHEWVITLRYMNIASH  
KQIHVYRGKPLEGFDAVPRIGASVTFYCAVLRFQPMGVPLNESVAIARSDKLR  
SLQLSRSGIGPITGFAHSPDIPLQIVNNGAPLIVKVLGSGTQIGVVLCTPOAA  
ESVIEAFMGLKQNIWQVEYIKAGGADIRCFVGVKVIASMKRQAKPGFGRSNLHGG  
NVASLKIITEERTAIRAAKVMGLSVAGVDILRSNHPGLVMEVNSSPGLEIEVTTGK  
NYAGMIIEHLEKNGPNQTRTKGK"  
complement (982..1602)  
/locus\_tag="PP0249"  
complement (982..1602)  
/locus\_tag="PP0249"  
/note="identified by Glimmer2; putative"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAN65881.1"  
/db\_xref="GI:24981608"

```
/translation="MSSSSAFCAIPRLPIILSIIFLYLHIRMQKVPALVKVHLIRRV
RFVTKTFLVIGLRWVALPDGLVAGLRKIDTGASTSSLHATEVEPFRDGGPWVR
FTHLGSVQHRRCRAPLVTMTKTIKSSNGHACTRVYIRFSLAGQWHEVEITLAC
RKNRYRLILGSKALIHQQLVNFGLKXVQKQFAPATLSPVTGAA"
1574. .1975
/gene="hs1R"
/locus tag="PP0250"
1574. .1975
/gene="hs1R"
/locus tag="PP0250"
/notes="identified by match to PFAM protein family HMM
PF01479"
/transl_table=1
/product="heat shock protein 15"
/protein_id="AAN65882.1"
/db_xref="GI:24981609"
/translation="MAQKAEDDDKVRLLDKMWAARFYKTRALAKAAIESGVHCGRGR
CKPGKPRVGDGFVIRTGDFERTVVVKALSVYRRGAPEAQALYEETESVRRREQAAE
MRKAGAMGVTTDGRPTKKQRQIHLHGSYD"
complement (2004. .2798)
/locus tag="PP0251"
complement (2004. .2798)
/locus tag="PP0251"
/notes="identified by match to PFAM protein family HMM
PF01569"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAN65883.1"
/db_xref="GI:24981610"
/translation="MDNRPLFOARWSWGPIAASLLPIALLCFWLPFGOILCLTDE
WLPESNLAPLADSTFVYITIGSLRPEDIVGVLAVLRIGDWPVKAQVRAFG
FLVTLVLRVIRALFESWDAAGQHKSPSIFDDVHLSDYFVNLRAEWELKDRSK
SFDGASVLLIHALFMSVSRGLAQIHWGLAALPRLVAGAHWGQDDYTGGLL
MAVLGWSYTTPLAAGSEALMRWTAPLKLSTLPLVGRMAVR"
2919. .3818
/locus tag="PP0252"
2919. .3818
/locus tag="PP0252"
/notes="identified by match to PFAM protein family HMM
PF01430"
/codon_start=1
/transl_table=1
/product="chaperonin, 33 kDa"
/protein_id="AAN65884.1"
/db_xref="GI:24981611"
/translation="MSDLPTDPTQRFIFDERIVRGESVLDSDSAAVLARHEYQPV
QVLLGELMAATALLVGAMKFDGLIILQARGAGPIPLIMVCSSEIREIRGMARYEADQI
PAGATLSQMPDGHLLTIDPVKQRYQGTVDLDGANLSCFTNYPVQSQOLNTRFNL
NAQGGKARGILLQOLPRDOPDDEERDSQHVVAKTLKPREWTEGNTLLHRLYH
EDAVRLDIQPLRPNCSRSRSGNALVSLGEHDAKALVDECGTVEDICQFCNERIF
FDASDVQLPAGGGTIDVASETRH"
3998. .5537
/gene="pckA"
/locus tag="PP0253"
/notes="This region contains an authentic frame shift and
is not the result of a sequencing artifact;
phosphoenolpyruvate carboxykinase, authentic frameshift;
similar to GB:U04815, GB:U04818, GB:U04824, GB:U04816,
GB:U04817, SP:P21127, PID:189481, PID:507158, PID:507160,
PID:507162, PID:507164, PID:507166, PID:507168,
PID:507427, and PID:507429; identified by sequence
similarity; putative"
5686. .6126
/locus tag="PP0254"
5686. .6126
/locus tag="PP0254"
/notes="identified by match to PFAM protein family HMM
PF03061"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein TIGR00051"
```

```
/protein_id="AAN65885.1"
/db_xref="GI:24981612"
/translation="WTEAPORTAFSHFPHILTRQDDNLDNLGHIAAGTVGHGFETAIQA
FLVQABDLRHGELAAVSSAADFYALPGFPDLLEVLGVTIRLAGSTVYRLALYR
PGSDACAAGTVQVQVFIERASGREVALPEALQVILSLGVLERQA"
complement (6183. .6665)
/locus tag="PP0255"
complement (6183. .6665)
/locus tag="PP0255"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAN65886.1"
/db_xref="GI:24981613"
/translation="MKKYISILLFSLLSGVMAGTTTGAIGGALGGVILGWSVWNS
IGSTGAGLGGAGGALGADKRGGEAAGGALGAAGNNVGRSVGGTTSYIG
AAAGGGAGGALGNYLGKADKDYRDGRRYRRGYDDDRHWRDGHYGRKHKHRWD
"
complement (6862. .9207)
/locus tag="PP0256"
complement (6862. .9207)
/locus tag="PP0256"
/notes="similar to SP:P07658; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="molybdopterin oxidoreductase, alpha subunit"
/protein_id="AAN65887.1"
/db_xref="GI:24981614"
/translation="MTSYQQLPONTPASPRYPKPHGAGGALRSVAKAWGSDNA
LKNIRALLQNQNGEDCPGCMGSDSPESRVFCEANGAKVWEATKRRVDAFFPAR
YVTSLLQSDIWELEYQUTPEMYPDAPSDRVDLPISDAFAFALABELANLPDQA
EFTYSGRASNAAYLYQLFVRYATNNPDCSNMCHSEASVGLSGVGVGQVTFDD
FEHADAIYWGQNPNTMPLDPLRDVAKRGQVVCINPLKRGELRFQHPQPLEM
LITNSDRTNTAFRPPALGDMALRGKAFVLOWERAQNGEPVDFHAFIABHGH
VDEYLAVDATPWSHIOAQSLTIADIELAAARMYCOGKRVIMCWAMGITOHRHSVPTI
QELVNLMLRGNIGVPGAGLCPRGHSNVQDRTMGINERPPALLDAIEKRGFPVP
RRGNHIVEALHMLGRKAFVIGLGNFAQATPDTERTAQRNCELSTVHISTKLMR
SHLVHGQALILFLGRITDLDQAGDQAVTVEDSFMSHASNGQLKPLSTQMRSEPA
VIAGIAAATLKGQFVDMHVLVADYRIDLIGDTIPGFSGNRSHDQVNTTYIGLDDRY
SRWATSTGRANFKANLLPDLDERVRASGLPDLIMQSMRSHDQVNTTYIGLDDRY
RGVQOREVLFAEADIIRLGFQPGOKVDIVSLWGDEHVRVQGFLLAIDIPAGQAA
AYVPEVNPVPLESIGVSGSHTPSKFIAIKLERAREDRIL"
complement (9213. .10043)
/locus tag="PP0257"
complement (9213. .10043)
/locus tag="PP0257"
/notes="similar to GB:X60787, GB:S41456, SP:Q01167,
PID:1388150, and PID:33854; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="formate dehydrogenase accessory protein FdhD"
/protein_id="AAN65888.1"
/db_xref="GI:24981615"
/translation="MNSKPEVCAASSPLSMFAASNTYVYVQLSDAKRDSFTFLAEVAL
AIYVGNLQAVMLVSPDLEDFAVGFSVGIVEGTAETIYDLKLSGSGSALYADLEIS
SRAFWNLKQRRQLAGTSGCGLCGVEALEQALPDVLVLPGLAPPLPAQWLVLGRERIDA
"
Query Match 4.7%; Score 45.4; DB 1; Length 300511;
Best Local Similarity 54.5%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 91; Conservative 0; Mismatches 76;

QY 645 GATCTTTGAAGAGTGCTCGGCTGACTTTTCGACCCCATCGATGCGGACCTTAT 704
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
31147 GCTGTTCCGGCGACCTGCATCGCGTTCGACCTGTGTCGTTTTCACGACGAAT 31088

QY 705 TCTCCATGACGAAGTTCGTGGGTGGCTTTTCCTTATGAGGCCGTGAAAAGTAAAGTCC 764
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
31087 GCAACACGACGACCATGTCTGCAAAATTTGGCCACATACACGCATCGAGGCTGC 31028
QY 765 GTTTGATGAGGCGGCATCGACCTGCNAGAGCTATTCTGTTACAC 811
```



```
KEGHQLOIKLILEVKKGBLLDVKVDNQIKNHLRLKVNIGVKTDYNIASLPFGY
IERKVPVANNQOYTSEMPDIDIRPLQSVTLTNEAESCTTFTQGVKEYQOQDOHLALT
LLATLTQGLKADLAYRGRASGDITKKGHVLITTEGRQALLGKHEFTLAIYVLGAPDE
HKTAERTKAFQANVSQROPFPHFLHRLDNKIQKTERKLGAKQTLGMLNPKRYLVS
ACPSYYTADKYIIRLENPTNTPOKLPLDPVSPFYNVAIEEVVANDNNTIIPAYGAVTL
RLDL"
3927. .3954
/ gene= "lin0449"
/ complement (3929. .4816)
/ gene= "lin0450"
/ complement (3929. .3956)
/ gene= "lin0450"
/ complement (3941. .4816)
/ gene= "lin0450"
/ notes= "similar to transcriptional regulator (LysR family)"
/ codon_start=1
/ transl_table=11
/ protein_id= "CAC95682.1"
/ db_xref= "GI:16412891"
/ db_xref= "GOA:Q92EK9"
/ db_xref= "SPTREMBL:Q92EK9"
/ translation= "MELRVLYNLYFLTVAREKTIISRAEVLHLSQPTLSKQLKELEELG
VTIFIRGNRSITTEDGLYLANRKEILSLVELTTSNLLQNETISGKIMIGAAETRGF
EFTGSVLHEUREXHPDITFQLHSGNADDDVLEKIEHGLLDFGLVINPAEKLOYEYLPIT
IEDTWGLVNTKRPDLAEKVVTPTSDIDANPIVSNQSFIDNQLASWLGHFQELNVIG
RYLLYNASLLAKENIASILCIDGIINTENTNLKFIPIFSPPLTAGISIVWKKNQTPSS
ASKEFLRLIOKTCSK"
4934. 5503
/ gene= "lin0451"
4934. 5503
/ gene= "lin0451"
/ notes= "similar to acetyltransferase"
/ codon_start=1
/ transl_table=11
/ protein_id= "CAC95683.1"
/ db_xref= "GI:16412892"
/ db_xref= "SPTREMBL:Q92EK8"
/ translation= "MTBQDLQKRLVQEOILQSGDLLDEIHEVKONNEOLIIELNTRYH
TKREVTKYLSIDITGKVPDPSVDVSLPSPDFGKHITFGKNLFINLNVTFVDLGLGITLE
DNVLIGPARLVTYVHLVSPKRGRLVAFICVKNAMIGANATILSGVITGENAIVA
ADATVTKDVPANVIVAGSPAKQIRKIIIE"
join (5503. .5508, 5517. .6263, 6272. .6302)
/ gene= "lin0452"
5503. 5508
/ gene= "lin0452"
5517. 6263
/ gene= "lin0452"
/ notes= "similar to oxidoreductase"
/ codon_start=1
/ transl_table=11
/ protein_id= "CAC95684.1"
/ db_xref= "GI:16412893"
/ db_xref= "GOA:Q92EK7"
/ db_xref= "SWISS-PROT:Q92EK7"
/ translation= "MITKNKVIITIGASSGTGEBATALLAEKGAKLVAARRVEKLEK
IVOTIKASGEALFMFTKNKVIITIGASSGTGEBATALLAEKGAKLVAARRVEKLEK
EDWEOMIDINIKGLNGLIAVLPSFIAQKSHLIATSSVAGLKAYPGGAYGATKWA
VRDLMEVLRMSAQEGNIRIATIPAAINTELLETITIDKTEQGMNLYKQYGVTFD
RIASIVAYAIDQPEDINNVNEFTVGPPTQPW"
6272. .6302
/ gene= "lin0452"
6362. .6884
/ gene= "lin0453"
6362. .6367
/ gene= "lin0453"
6375. .6884
/ gene= "lin0453"
/ codon_start=1
/ transl_table=11
/ protein_id= "CAC95685.1"
/ db_xref= "GI:16412894"
/ db_xref= "SPTREMBL:Q92EK6"
/ translation= "MNRKQLEQLDKCLYEYVDEQETNDFRVGYIRAFIDFSFIILEK
```

```
YHTRVGFDSGFALVANEALFKVQODTOYLKPYQNLNRPOEQIPELSEDDGDTLLNPLEWT
KAKKLVEIMTVWDELLIGTIEETIEENLAVMLNTVAPEEFQSDGKCFIECDNIIVDICID
TBLKRFITTT"
7123. .13641
/ gene= "lin0454"
7123. 7128
/ gene= "lin0454"
7138. .13641
/ gene= "lin0454"
/ notes= "similar to cell wall-associated protein precursor
wapa (B. subtilis)"
/ codon_start=1
/ transl_table=11
/ protein_id= "CAC95686.1"
/ db_xref= "GI:16412895"
/ db_xref= "GOA:Q92EK5"
/ db_xref= "SPTREMBL:Q92EK5"
/ translation= "MKIHAKARVLSLIAFMFLFSLPGYLAEDSPVGVADPDKK
AGEQAPTEVEKEERTENEVFDNHDGSPFKQIFADPINMEVDGMKRIIDANVEADSD
MIVPKOTPLELGFLEKMEVDGAYOKLTKAGAEVTFRLKAGTGENEQAVTQPAKYEN
EVTYENVPPTDLRHLTFPOSVKEDLVLRPNQVDTVYVOIETKLDELAEAGDVLFEK
NKSDETMYTILPKPMTDSNVGTETGERAAQSENVSFKQLTKTYLQELQKVDREWLND
DAREYPIYDPSVRLDKINYANINSAPKTETNIGSKLWDSGQNAVTLTKGKWNSTGN
NAAPLKMDSITLNKATISKATLKVYNIMHSPVTKNDLWYENAKMSPOQVWNTVTP
PVTRLGSVNVGRGEWANLDVTKTQAWASGRVNVNGFRLGTNINKNYKKVYASNNK
NYPYLEVNYTYAQPEKPTVKTSSGVTGTGMDLSGKAVPGATSYNIVISNGYKVE
FNTKSTATWTSTGKKI FPTNDRIAKGEREFHDKGTDFALDPRQAYENAFQAGSTF
GLNLRYLFRVQAVPFGSGSPNSDLVFAYPEIKPQAPVAKAYSNLAKHETGYVELN
WEKSPWADGKYLVFNGAEYEDVGAETKWTGKNGIWEPTKEIABGKALHHDHKG
AEALAKOPSPVYTNSSGGYKERTNYFRVIAIQAGNNNTSVQSEPAFSPPEALNKQL
GMVDYMTSPVRGVEENATNGFLFHETDFNLEGRGFSINVRTFNSQDDAGYLKIEE
WTSITLEKLVVEEENGINVWVESDKIHRFTKGDKEYEAPGIYSEITKNADGVLKIEE
DKSETPLVDGRLLKSEKDTGKNELTVEYTDGKLTSLRDSASRTVTLTYEGELAKELVG
PEDRKISYTNQDKELLSSSTARGKLYRYGTGTLTSLTSDYDPKHTKEKPYETTPAYEE
EKLTETIDPVGKKITLSDYKAEQOQTLTNEKKKTIYSYNDAGNPKKEIWDADGLKT
TITYESNNLVKEVNPKEQTEYAYDAGNITKATDAYTESYTYNDNNDVTSSTDTE
GRKTYTIDGADAVSETLATESQVSVTYQDAYGNPIRSGSELSGGLNLSQSFYEG
AGVSNWTLIOSDAKGSMTFNTOSAPGALGGSGSVKLTSEANSTVKGYSSVTVRDVE
PETTYTFSAMIKTSGMTNADALLIGRLODANAKDVTDAGVQSNRATSIKNGDWMVKR
QLTFKSKNTRQVLLVLDNEQAPAPHGKGIWYDNVQFEGKSVASSNPVNVNRSFENH
NGTLPTGMKMTGNALTQAKVDNDSHSGDSAVYFERKATSEATHTIVQDPVNVQKSA
KALTI SALSSEDAKANGSVATMSNDYSVMGTVYYODGTTSSVGQGFPLGTNDMNRSA
VVVQPTPKVMKIKVYTFMRNGLTKAWDFDVRVEGBVLTKNEDYDAGSNVYVTAZYDSE
GRKISFTYDIYGNKTSSETDEKGNKTLTYDADNALIDTKLANGTSVAYKYDDNGNTPE
XNVTASGKTQKNIYEYVDVNKITAFTDALNRITIKYEYDAAGNETKALMPNGRVTESTY
```

Query Match 4.0%; Score 38.6; DB 1; Length 249050;  
Best Local Similarity 61.4%; Pred. No. 16;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy	453	GCAGAGCGGAGCTCCAGGTCGGATGATCTCTGTGATCTCCCTGGTGGCGAGGAATT	512
Db	147936	GCAGCGCAAAATTGGCAGCGTCGATGAAATTTATCAAGTTGGCGTGGCAGCGGAATC	147877
Qy	513	TCGTCTCCCTCCCTCCCGGCGAGAATCCATTGCTGGAAGAAGA	553
Db	147876	GCTGCCCTACTACGTACAGCATCCATCTCTAAAGTAGA	147836

RESULT 8  
AX417038/c

LOCUS	AX417038	349980 bp	DNA	linear	PAT 02-SEP-2002
DEFINITION	Sequence	4029 from Patent WO0228891.			
ACCESSION	AX417038				
VERSION	AX417038.1	GI:21449648			
KEYWORDS					
SOURCE	1				
ORGANISM	1				
REFERENCE	1				
AUTHORS	Kunst, F. and Glaser, P.				
TITLE	Listeria innocua, genome and applications				
JOURNAL	Patent: WO 0228891-A 4029 11-APR-2002;				





```

/note="assembly_fragment
clone_end:SP6
vector_side:left"
1295..2852
/note="assembly_fragment"
2953..5751
/note="assembly_fragment"
5852..9886
/note="assembly_fragment"
9987..26134
/note="assembly_fragment"
26235..36673
/note="assembly_fragment"
36774..50148
/note="assembly_fragment"
50249..83726
/note="assembly_fragment"
83827..168186
/note="assembly_fragment
clone_end:T7
vector_side:right"

ORIGIN
Query Match 4.0%; Score 38.4; DB 2; Length 168186;
Best Local Similarity 57.5%; Pred. No. 17;
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 749 TGAAAGTAAGTCCGTTTGATGAGGCGGCATCGAGCGTCAAGACCTATTCGTTA 808
DB 38835 TTAATATGAGGAGCATATGATGAGGAGGATTCATCGGCGAGTCAAGACCTATTTGT 38776

QY 809 CACCTTCAGCGGTCTAAATTTTGTAACTGCTTTGAAATCGTGAGTCGATCGGTCACT 868
DB 38775 TTGCTTTGATGATATTTATTTTGAATCTGTTTTTAACTTGAATGATGATGTCACGT 38716

RESULT 13
AC130283/c
LOCUS
DEFINITION Mus musculus clone RP23-136E18, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
ACCESSION AC130283
VERSION AC130283.2 GI:28629906
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 177306)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-136E18
Unpublished

2 (bases 1 to 177306)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,V., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,J., Menga,V.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenka,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (09-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 177306)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,V.,
Collymore,A., Cooke,P., Corum,B., Dorr,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenka,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 2, 2003 this sequence version replaced gi:22165200.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27038
Center clone name: 136_E18
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 175079 bases at least Q40
Consensus quality: 176013 bases at least Q30
Consensus quality: 176308 bases at least Q20
Insert size: 218000; agarose-fp
Insert size: 176706; sum-of-contigs
Quality coverage: 6.7 in Q20 bases; agarose-fp
Quality coverage: 8.2 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 65842: contig of 65842 bp in length
* 65843 65942: gap of 100 bp
* 65943 68847: contig of 2905 bp in length
* 68848 68947: gap of 100 bp
* 68948 73881: contig of 4934 bp in length
* 73882 73981: gap of 100 bp
* 73982 81087: contig of 7106 bp in length
* 81088 81187: gap of 100 bp
* 81188 108661: contig of 27474 bp in length
* 108662 108761: gap of 100 bp
* 108762 140549: contig of 31788 bp in length
* 140550 140649: gap of 100 bp
* 140650 177306: contig of 36657 bp in length.
*
* Location/Qualifiers
* ..177306
* /organism="Mus musculus"

```

REFERENCE  
AUTHORSTITLE  
JOURNAL  
COMMENTFEATURES  
source

```

/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-136E18"
/clone_lib="RPI-23 Female Mouse BAC"
1..65842
/notes="assembly_fragment"
vector_side:left
65943..68847
/notes="assembly_fragment"
68948..73881
/notes="assembly_fragment"
73982..81087
/notes="assembly_fragment"
81188..108661
/notes="assembly_fragment"
108762..140549
/notes="assembly_fragment"
140650..177306
/notes="assembly_fragment"
vector_side:right"

ORIGIN
Query Match      4.0%; Score 38.4; DB 2; Length 177306;
Best Local Similarity 57.5%; Pred. No. 18;
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 749 TGAAGTAAAGTCGCTTTGATGAGGAGCGGATCGCAGCTGCAAGACTATTCTGTTA 808
Db 15161 TTAATATGGAGGATCATATGAAGAGAGAGATTCATGCGCAGCATCCAGACCTATTTTGT 15102

QY 809 CACCTTCAGCGGCTCTAAATTTTGTAACTGCTTTGAAAATCTGTAGCTGATCGTCAGT 868
Db 15101 TTGCTTTGATGATTTATTTTGAATCTGTTTTTTAACTTGAATGATGTCAGT 15042

RESULT 14
AC132844/c
LOCUS          181360 bp      DNA      linear      HTG 16-FEB-2003
DEFINITION    Mus musculus clone RP24-481P13, WORKING DRAFT SEQUENCE, 3 unordered
               pieces.
ACCESSION     AC132844
VERSION       AC132844.3 GI:28394960
KEYWORDS      HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE        Mus musculus (house mouse)
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 181360)
AUTHORS       Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
               Barna,N., Bastien,V., Bloom,T., Boguslavkiy,I., Boukhgalter,B.,
               Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
               Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
               Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
               Garmhauser,L., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
               Karatas,A., Kells,C., Landers,T., Levine,R., Linderblad-Toh,K.,
               Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
               McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J.,
               O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
               Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
               Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,
               Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
               Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
               Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
               Zembek,L., Zimmer,A. and Zody,M.
               Direct Submission

TITLE

```

## JOURNAL

REFERENCE  
AUTHORS

Submitted (04-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 181360)

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N., Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,I., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y., Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K., Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Garmhauser,L., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Linderblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C., Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (16-FEB-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 16, 2003 this sequence version replaced gi:25956358.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L27224

Center clone name: 481.P.13

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 180535 bases at least Q40

Consensus quality: 180783 bases at least Q30

Consensus quality: 180973 bases at least Q20

Insert size: 181160; sum-of-contigs

Quality coverage: 11.7 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 3 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 103657: contig of 103657 bp in length

\* 103658 103757: gap of 100 bp

\* 103758 110625: contig of 6868 bp in length

\* 110626 110725: gap of 100 bp

\* 110726 181360: contig of 70635 bp in length.

Location/Qualifiers

1..181360

/organism="Mus musculus"

/mol\_type="genomic DNA"

/db\_xref="taxon:10090"

/clone="RP24-481P13"

/clone\_lib="RPI-24 Male Mouse BAC"

1..103657

/note="assembly\_fragment"

clone end:SP6

vector side:left"

103758..110625

misc\_feature

misc\_feature

## TITLE

## JOURNAL

## COMMENT

```

/misc_feature
110726..181360
/Note="assembly_fragment
clone_end:T7
vector_side:right"

```

## ORIGIN

```

Query Match      4.0%; Score 38.4; DB 2; Length 181360;
Best Local Similarity 57.5%; Pred. No. 18;
Matches 69; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 749 TCAAAAGTAAAGGTCGCTTTCATGAGGAGCGCATCGCAGCTGCAAGAGCTATTTCGTTA 808
Db 15169 TTAATATGAGGAGATCATATGAGAGGAGATTCATCGCAGCATCAAGACCTATTTTGT 15110

Qy 809 CACCTTACGCGGTCCTAATTTTGTAACTGCTTTGAAATCGTGAGTCGATCGCTCAGT 868
Db 15109 TTGCCTTGTATGATTTTATTTTGAATCTGTTTTTTAACTTGAATGTATGTCCTCAAGT 15050

```

## RESULT 15

```

AC129256
LOCUS      Rattus norvegicus clone CH230-384015, WORKING DRAFT SEQUENCE.
DEFINITION
ACCESSION  AC129256
VERSION    AC129256.4 GI:25072556
KEYWORDS   HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE     Rattus norvegicus
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus

```

## REFERENCE

```

1 (bases 1 to 214511)
Munzy,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Albrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshueh,L., Louibeged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Nwaokemelehu,O., Okunolu,G., Olarnpusagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quirao,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

```

## COMMENT

```

Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 214511)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (28-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GTGP
Center clone name: CH230-384015
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 201412 bases at least Q40
Consensus quality: 203652 bases at least Q30
Consensus quality: 205020 bases at least Q20
Estimated insert size: 208107; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
  (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
  consists of 1 contigs. Gaps between the contigs
  are represented as runs of N. The order of the pieces
  is believed to be correct as given, however the sizes
  of the gaps between them are based on estimates that have
  been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 214511: contig of 214511 bp in length.
  Location/Qualifiers
    1..214511
      /organism="Rattus norvegicus"
      /mol_type="genomic DNA"
      /db_xref="taxon:10116"
      /clone="CH230-384015"
  misc_feature
    1..1753
      /note="wgs_end_extension
clone_end:Sp6"

```

## FEATURES

```

source
misc_feature

```

```

misc_feature      3142..4830
                  /note="wgs_end_extension
clone_end:Sp6"
6595..7733
                  /note="wgs_end_extension
clone_end:Sp6"
8697..9302
                  /note="clone_boundary
clone_end:Sp6
site:
end_sequence:BZ281669"
misc_feature      complement(207915..208580)
                  /note="clone_boundary
clone_end:T7
site:
end_sequence:BZ281668"
211140..214511
misc_feature      /note="wgs_end_extension
clone_end:T7"

```

ORIGIN

```

Query Match      3.9%; Score 38; DB 2; Length 214511;
Best Local Similarity 47.8%; Pred.No. 23;
Matches 110; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY      513 TCTGCTCTCGCCTGCCGGCAGAAATCCATTGCTGAAGAAGATGCTCGAAGAGTTTGTGCC 572
Db      136957 TCTGGAGTAGAGAGAAGGTGTGCAGAAAGTGGGGGAGAGCTGGGCTTCAGCTAGGGCTG 137016

QY      573 GCGATTTGCACCTCGCTCGACGGTGTCTACTCTCGGGGATACCTCGTGGAAAGCATTCCT 632
Db      137017 GAGTGTGGGAGTTGCTCTGGGCTGGCATCCTACTGAGATACTGCTGGAACCTATTCCAGT 137076

QY      633 ATTGGAACGAGAGATCTTTGNAGAGGTGCTCGGCTGACTTTCACCCCATGTCGAAT 692
Db      137077 CTCAGAAGGAAGTTTCCAGGAGATCTGGATGTCCTGTGTGGATGCTACTAGACTGAG 137136

QY      693 GCCGGACCTTATCTCCATGACGAAGTTCGTGGGTGGCTTTTCCTTATGG 742
Db      137137 GCCAGAGATGTCCTGTCTTCCTATCGAGTACCTGATGCCCATCCTCCAGG 137186

```

Search completed: October 2, 2004, 03:13:40  
Job time : 3966.41 secs

P Lank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 19:21:23 ; Search time 416.286 Seconds  
(without alignments)  
9919.257 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtgaacagcagtcagcgcat.....ttgggcgcgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseqn\_29Jan04.\*

1: Geneseqn1990s.\*

2: Geneseqn1990s.\*

3: Geneseqn2000s.\*

4: Geneseqn2001as.\*

5: Geneseqn2001bs.\*

6: Geneseqn2002s.\*

7: Geneseqn2003as.\*

8: Geneseqn2003bs.\*

9: Geneseqn2003cs.\*

10: Geneseqn2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.6	4.0	2000	7 ADA71938	Ada71938 Rice gene
C 2	38.6	4.0	110000	6 ABQ69245_05	Continuation (6 of
C 3	38.6	4.0	110000	6 ABQ69245_06	Continuation (7 of
C 4	38.6	4.0	110000	6 ABQ67197_05	Continuation (6 of
C 5	38.4	4.0	2000	7 ADA71938	Ada71938 Rice gene
C 6	37.8	3.9	1413	4 ABL24039	Ab124039 Drosophil
C 7	35	3.6	1431	4 ABL07657	Ab107657 Drosophil
C 8	35	3.6	4200	4 ABL07656	Ab107656 Drosophil
C 9	34.8	3.6	1590	2 AAV19125	Aav19125 Nucleotid
C 10	34.8	3.6	1590	2 AAX25066	Aax25066 Brassica
C 11	34.8	3.6	1590	6 ABA97357	Aba97357 Brassica
C 12	33.8	3.5	7133	6 ABL58181	Ab158181 Murine St
C 13	33.6	3.5	1073	3 AAC39454	Aac39454 Arabidops
C 14	33.2	3.4	3780	4 ABL24038	Ab124038 Drosophil
C 15	33	3.4	348	6 ABL81387	Ab181387 Human ova
C 16	33	3.4	391	3 AAC09571	Aac09571 Human sec
C 17	33	3.4	481	4 AAL14686	Aal14686 Human bre
C 18	33	3.4	482	4 AAL23548	Aal23548 Human bre
C 19	33	3.4	493	4 AAS27046	Aas27046 cDNA enco
C 20	33	3.4	493	9 ADB93224	Adb93224 Human cDN
C 21	33	3.4	494	6 ABL79833	Ab179833 Human ova
C 22	33	3.4	532	3 AAC99042	Aac99042 Human pan
C 23	33	3.4	532	4 AAF72769	Aaf72769 Human pro

C 24	33	3.4	579	3 AAC77387	Aac77387 Human ORF
C 25	33	3.4	918	3 AAC76648	Aac76648 Human ORF
C 26	33	3.4	1022	6 ABK46152	Abk46152 cDNA enco
C 27	33	3.4	1097	3 AAA96893	Aaa96893 Nucleotid
C 28	33	3.4	1102	4 AAI58936	Aai58936 Human pol
C 29	33	3.4	1102	8 ADB48918	Adb48918 Novel hum
C 30	33	3.4	1129	3 AAC77841	Aac77841 Human can
C 31	33	3.4	1129	4 AAF72745	Aaf72745 Human pro
C 32	33	3.4	1129	4 AAF72770	Aaf72770 Human pro
C 33	33	3.4	1224	4 AAI60722	Aai60722 Human pol
C 34	32.8	3.4	7357	4 ABL18274	Ab118274 Drosophil
C 35	32.6	3.4	103599	4 ABX04971	Abx04971 S. cinnam
C 36	32.4	3.3	588	4 ABL25387	Ab125387 Drosophil
C 37	32.4	3.3	1167	6 RAD46542	Rad46542 Mouse dia
C 38	32.4	3.3	1881	6 ABK99846	Abk99846 Babesia c
C 39	32.4	3.3	2262	8 AAD56889	Aad56889 Mouse dia
C 40	32.4	3.3	4188	7 ABX13487	Abx13487 B. popill
C 41	32.4	3.3	8294	6 ABL34430	Ab134430 Human imm
C 42	32.2	3.3	345	5 ABV45947	Abv45947 Human pro
C 43	32.2	3.3	1707	4 ABL08339	Ab108339 Drosophil
C 44	32.2	3.3	6594	4 ABL08338	Ab108338 Drosophil
C 45	32.2	3.3	10982	4 AAS46264	Aas46264 DNA enco

#### ALIGNMENTS

##### RESULT 1

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX

AC ADA71938;

XX

DT 20-NOV-2003 (first entry)

XX

DE Rice gene, SEQ ID 5263.

XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

XX

OS Oryza sativa.

XX

PN W02003000898-A1.

XX

PD 03-JAN-2003.

XX

PF 22-JUN-2001; 2001WO-IB001105.

XX

PR 22-JUN-2001; 2001WO-IB001105.

XX

PA (SYGN ) SYNGENTA PARTICIPATIONS AG.

XX

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX

PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX

DR WPI; 2003-175250/17.

XX

PT Identifying at least one gene involved in plant resistance or response to

XX

PT pathogenic infection for conferring resistance or tolerance to a plant to

XX

PT bacterial, fungal or viral infection by determining or detecting plant

XX

PS Claim 27; SEQ ID NO 5263; 899pp; English.

XX

CC The present invention relates to a method (M1) for identifying genes

CC

CC involved in plant resistance or response to pathogenic infection. M1

CC

CC comprises identifying a gene whose expression is significantly altered in

CC

CC the incompatible interaction of plant gene expression relative to

CC

CC expression of the gene in an uninfected plant, in a mutant plant that

CC

CC does not express a gene associated with response to pathogenic infection,

CC

CC or in a corresponding incompatible or compatible interaction. (M1) is

CC

CC useful for conferring resistance to resistance or tolerance to a plant to

CC

CC bacterial, fungal or viral infection. The present sequence was used to

WP ABQ671  
WP ABQ671  
WP ABQ671

WP	ABQ67197_01	100001	210000
WP <td>ABQ67197_02 <td>200001 <td>310000 </td></td></td>	ABQ67197_02 <td>200001 <td>310000 </td></td>	200001 <td>310000 </td>	310000
WP <td>ABQ67197_03 <td>300001 <td>410000 </td></td></td>	ABQ67197_03 <td>300001 <td>410000 </td></td>	300001 <td>410000 </td>	410000

WP ABQ67197\_04 400001 510000  
WP ABQ67197\_05 500001 610000  
WP ABQ67197\_06 600001 710000  
WP ABQ67197\_07 700001 810000  
WP ABQ67197\_08 800001 910000  
WP ABQ67197\_09 900001 1010000  
WP ABQ67197\_10 1000001 1110000  
WP ABQ67197\_11 1100001 1163020

Query Match 4.0%; Score 38.6; DB 6; Length 110000;  
Best Local Similarity 61.4%; Pred. No. 1.5;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 453 GCAGCGGAAGCTCTCCAGGCTGCGATGATCTCTGTCGATCTCCTCGTGGCGAGGAATT 512  
DB 13671 GCAGCGCAATTGGCAGCGTGCATGAATTTATCAAGTTGCGGTGCGCAGGAATC 13612

QY 513 TCTGCTCTCGCTGCGCGGAGCAATCATGCTGAAGAAGA 553  
DB 13611 GCTGCCCTAGCTTACGCTACAGCATCCATCTCTAAAGTAGA 13571

RESULT 5  
ADA71938  
ID ADA71938 standard; DNA; 2000 BP.  
AC ADA71938;  
DT 20-NOV-2003 (first entry)  
XX Rice gene, SEQ ID 5263.  
DE Plant; bacterial infection; fungal infection; viral infection; rice;  
KW gene; ds.  
XX Oryza sativa.  
XX WO2003000898-A1.  
XX 03-JAN-2003.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX 22-JUN-2001; 2001WO-IB001105.  
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.  
XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;  
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;  
XX WPI; 2003-175290/17.  
XX Identifying at least one gene involved in plant resistance or response to  
PT pathogenic infection for conferring resistance or tolerance to a plant to  
PT bacterial, fungal or viral infection by determining or detecting plant  
PT gene expression.  
XX Claim 27; SEQ ID NO 5263; 899pp; English.  
XX The present invention relates to a method (M1) for identifying genes  
CC involved in plant resistance or response to pathogenic infection. M1  
CC comprises identifying a gene whose expression is significantly altered in  
CC the incompatible interaction of plant gene expression relative to  
CC expression of the gene in an uninfected plant, in a mutant plant that  
CC does not express a gene associated with response to pathogenic infection,  
CC or in a corresponding incompatible or compatible interaction. (M1) is  
CC useful for conferring resistance to resistance or tolerance to a plant to  
CC bacterial, fungal or viral infection. The present sequence was used to  
CC illustrate the invention.

SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
XX Query Match 4.0%; Score 38.4; DB 7; Length 2000;

Best Local Similarity 10.7%; Pred. No. 0.21;  
Matches 42; Conservative 173; Mismatches 179; Indels 0; Gaps 0;

QY 6 CAGCAGTACGGCATCGACGAGAGTGGAGCATCGATCTCGGGGGCGCTGTCTATAA 65  
DB 35 SRMSRKMGSMSKYRKCSCGCKMTTTRKSKWYSASSASGRTGSKWSSGSYSKGKMKRY 94  
QY 66 GCGTTTGGGTTTGACGCGCAACGATATAACGTCCTGAGCGCTGTGACATGTCTGGCGCT 125  
DB 95 KRKRWRGR 154  
QY 126 TGCCGGTTTCAAGCCGGGAGATCGTGGGTGACTCGACCACTCCACGCTTGGCGTTCA 185  
DB 155 SCREKSKSGSWKTCERGARGSGSWSSGAKYSGSKMWMSSCGRSGGRRSAISRY 214  
QY 186 GAGATCATGATGCTGGTTCGGCGAGCATGGGCAAGCCGATCGCCACCGAAGTCGAGA 245  
DB 215 YGTSRKYGTYKMTYYSASRCMEAYMTTSYWCSSYTWCRSKRRSMNMWKRMSRS 274  
QY 246 AGATTTCCGTAAGAAGACGCTTCGGCAGTGGGTGATAATGGCTTCGCCGCTACTTATGC 305  
DB 275 YGWYSWYKMMCTAYKSYYSRWCMYRGGGWRGATRYWGRGYMSRMAMMYKKMYWYRG 334  
QY 306 GGACAATTTAAACATCGCCACGAACTCGCAGCTCAACGAGTACTGTCTGTGACGAAGC 365  
DB 335 YKGMKRGWAGRMMSMRSMCRWSKACYMWRMWRMWRMWRMWRMWRMWRMWRMWRM 394  
QY 366 ATTACAGCGCTAAGGCGCATATGGAACGGAAGGC 399  
DB 395 RYKRMRGYSRMSRCKRARMKRCRGRAMKGC 428

RESULT 6  
ABL24039/c  
ID ABL24039 standard; DNA; 1413 BP.  
XX ABL24039;  
AC ABL24039;  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23590.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ds.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX Claim 1; SEQ ID NO 23590; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1413 BP; 353 A; 434 C; 404 G; 222 T; 0 U; 0 Other;  
Query Match 3.9%; Score 37.8; DB 4; Length 1413;  
Best Local Similarity 45.0%; Pred. No. 0.28;  
Matches 141; Conservative 0; Mismatches 172; Indels 0; Gaps 0;  
QY 425 TTGATGAGCATCGAAGCGCTTAAAGCGGAGCGAGCTCTCAGGCTCGGATGATCT 484  
DB 653 TTAGGGAGTCCAGCTCTTGGCATAGCTAGCTTCAATGGTCTCAGGCCGTGATCCACC 594  
QY 485 CTGTGCTCTCCCTGGTGGCGAGGAATTTCTGCTCTCCCTGCCGGGAGGAATCCATTGC 544  
DB 593 AGCCGGGAGGACAGGGTGTGATGATCTTCTGCTGGCGCTGGATGGTCAACTCCCGATGG 534  
QY 545 TGAAGAAGATGTCGAAGAGTTTGGCGCGAATTTGCACCTCGCTCGACGGTGTCTTACC 604  
DB 533 TTCAATAGACTGGCTATCTGCTTCTGCTTCGCTTCAGTCGGGATTCGAAGAGCAAGAGC 474  
QY 605 TCGGGGATCTCGTGAAGCATTCCTATTTCGAACGAGAGATCTTTGAAGAGGTGCTCG 664  
DB 473 TCGGAGGTCAACGAGTTCTATCTGATCAGATTTCTTCTGTAACGATAGTTGCTGG 414  
QY 665 GCCTGACTTTGACCCCATGCTCGAATGCCGACCTTATTCTTCATCAACGAAGTTCTGTG 724  
DB 413 ACCTGATGATACGCTTGGCCACTTCATGATGTGCTCGTCTGGAAGCGCACCTTG 354  
QY 725 GGTGGCTTTTCT 737  
DB 353 GAGCGCAGCTCT 341

RESULT 7  
ABL07657  
ID ABL07657 standard; cDNA; 1431 BP.  
AC ABL07657;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 17453.  
DE Drosophila melanogaster expressed polynucleotide; insecticide;  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PSDB; ABB63554.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.  
XX Claim 1; SEQ ID NO 17453; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 1431 BP; 390 A; 341 C; 342 G; 358 T; 0 U; 0 Other;  
Query Match 3.6%; Score 35; DB 4; Length 1431;  
Best Local Similarity 56.5%; Pred. No. 2.3;  
Matches 65; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 825 AATTTTGTAACTGCTTTGAAATCGTGAATCGATCGTCAGTGGCTCCCTGACCTGGC 884  
DB 942 AATTTTGTGAACGACTTTCAAAATCCTCAGGCATTTGGTAGATATCTCAATAAGTTGGC 1001  
QY 885 TTGGAAACTGAGCGGTGGGTAGCGGAGATCCAGACCATCTGATTCACCTTAAC 939  
DB 1002 GGATAACAAAAAGCTGTACAATTCGTACCGCCACACAACTAAATCGCGGTAAC 1056  
RESULT 8  
ABL07656/c  
ID ABL07656 standard; cDNA; 4200 BP.  
XX ABL07656;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 17450.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical; gene; ss.  
XX Drosophila melanogaster.  
OS  
XX WO200171042-A2.  
XX 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US009231.  
XX 23-MAR-2000; 2000US-0191637P.  
XX 11-JUL-2000; 2000US-00614150.  
XX (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI WPI; 2001-656860/75.  
XX P-PSDB; ABB63553.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions.  
XX Claim 1; SEQ ID NO 17450; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly



Db 1120 GGTCCGGCGCGTCTACATCAACGAGCGCTCCGAAAGGAGCCTGTCGGCGCTTACG 1061  
QY 230 CCACCGAAGTCGAGAGATTTCGTAAGAGAGCGCTTCGCGAGTGGTTCATATGGCT 289  
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCCAAAGATGCTTCAGAGTGCATCACCACGGTG 1001  
QY 290 TCGCGCGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985  
RESULT 11  
ABA97357/c  
ID ABA97357 standard; cDNA; 1590 BP.  
XX AC ABA97357;  
XX  
XX 12-APR-2002 (first entry)  
XX Brassica napus beta cyclase cDNA clone.  
XX  
XX Carotenoid compound; oilseed; plant; vitamin A; ss; gene.  
XX  
XX Brassica napus.  
XX  
XX WO200188169-A2.  
XX  
XX 22-NOV-2001.  
XX  
XX 11-MAY-2001; 2001WO-US015264.  
XX  
XX 12-MAY-2000; 2000US-00570140.  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
XX Shewmaker CK;  
XX  
XX WPI; 2002-075318/10.  
XX  
XX Altering carotenoid level/content in oilseed plants, by transforming  
XX plants with construct having DNA sequence encoding a carotenoid  
XX biosynthesis gene, plastid transit peptide, transcription initiator and  
XX terminator.  
XX  
XX Disclosure; Page 72-73; 114pp; English.  
XX  
XX This invention relates to altering the carotenoid content or composition  
XX in seed from a host plant. This is achieved by transforming cells of host  
XX plant with a construct of operably linked components, a transcriptional  
XX initiation region from a gene expressed in a plant seed, a plastid  
XX transit peptide, a nucleic acid sequence encoding a carotenoid  
XX biosynthesis gene from eukaryotic source, and a transcriptional  
XX termination region. The method is useful for increasing carotenoid  
XX content in seed and endosperm of a host plant and altering carotenoid  
XX composition in a plant seed. The method is also useful for screening  
XX transformed corn seeds or transformed endosperms, where the transformed  
XX seed and endosperms are visually determined and selected based on yellow,  
XX orange or red colour as result of the increased carotenoid content. The  
XX transformed seeds provide a source of modified oils and the oil extracted  
XX from the seeds is useful as a food colourant, or as a food oil with high  
XX alpha and beta-carotene levels for prevention of vitamin A deficiency  
XX which can result in night blindness. The level of lutein is also  
XX increased in seeds of transformed plants. The modified oil obtained from  
XX the transformed seeds are more oxidatively stable than naturally  
XX occurring oils and are low saturate, high oleic and low linolenic. This  
XX sequence represents the nucleotide sequence of the Brassica napus beta  
XX cyclase cDNA clone  
XX  
SQ Sequence 1590 BP; 438 A; 440 C; 352 G; 360 T; 0 U; 0 Other;  
Query Match 3.6%; Score 34.8; DB 6; Length 1590;  
Best Local Similarity 52.0%; Pred. No. 2.9;  
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

QY 112 ACATTGCTCGCGCTTGGCGTTTGAAGCCGGGAGATCGCTGGTTGACTCGACCACTCCA 171  
Db 1180 ACGAGTTTGGTTGACGAGTTTCGAGCCATGGACTTGTCTAGACTGCTCGACACCACT 1121  
QY 172 CG--CCTTGGCGTTTCAGAGATCATGACTGCTCGGGGAGCANTGGGCCAAGCGGTACG 229  
Db 1120 GGTCCGGCGCGTCTACATCAACGAGCGCTCCGAAAGGAGCCTGTCGGCGCTTACG 1061  
QY 230 CCACCGAAGTCGAGAGATTTCGTAAGAGAGCGCTTCGCGAGTGGTTCATATGGCT 289  
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCCAAAGATGCTTCAGAGTGCATCACCACGGTG 1001  
QY 290 TCGCGCGTACTTAATGC 305  
Db 1000 TTAGATTCATCAGGC 985  
RESULT 12  
ABL58181/c  
ID ABL58181 standard; DNA; 7133 BP.  
XX AC ABL58181;  
XX  
XX 05-AUG-2002 (first entry)  
XX Murine Stable Tubulin Only Polypeptide, STOP, gene fragment #1.  
XX  
XX STOP; neuroleptic; tranquiliser; antidepressant; transgenic animal;  
XX Stable Tubulin Only Polypeptide; drug screening; schizophrenia; gene;  
XX schizoaffective disorder; anxiety; paranoia; depression; murine; ds.  
XX  
XX Mus musculus.  
XX  
XX WO200241691-A2.  
XX  
XX 30-MAY-2002.  
XX  
XX 23-NOV-2001; 2001WO-PR003701.  
XX  
XX 24-NOV-2000; 2000FR-00015240.  
XX  
XX (COMS ) COMMISSARIAT ENERGIE ATOMIQUE.  
XX (INRM ) INST NAT SANTE & RECH MEDICALE.  
XX  
XX Andrieux A, Job D, Denarier E, Bosc C, Vernet M;  
XX  
XX WPI; 2002-463598/49.  
XX  
XX Recombinant mammal useful in screening for agents to treat e.g.  
XX schizophrenia, comprises a modified allele for stable tubulin only  
XX polypeptide (STOP) protein.  
XX  
XX Example 1; Fig 2; 57pp; French.  
XX  
XX The present invention relates to a non-human recombinant animal (A),  
XX which contains a modified allele of the gene for STOP (Stable Tubulin  
XX Only Polypeptide) protein. (A), or their cell extracts or organ slices  
XX (preferably neuronal or cerebral) are used to screen for drugs useful in  
XX treatment of schizophrenia or schizoaffective disorders that have an  
XX anxiety, paranoia or depressive component. The present sequence is a gene  
XX fragment of the murine STOP gene, used to illustrate the invention  
XX  
XX Sequence 7133 BP; 1765 A; 1941 C; 1718 G; 1709 T; 0 U; 0 Other;  
XX  
XX Query Match 3.5%; Score 33.8; DB 6; Length 7133;  
XX Best Local Similarity 51.7%; Pred. No. 14;  
XX Matches 77; Conservative 0; Mismatches 72; Indels 0; Gaps 0;  
QY 132 TTTGAAGCCGGGAGATCGCTGGTTCGACTCGACCACTCCACGCTTGGCGTTTCAGAGAT 191  
Db 4535 TTTCTGTCGCCGAGCACCACCGCGCTCTCTCCAGCACCTCCGGGTCCCGGGCATCAGCAGAA 4476

```
Qy 192 CATGGACTGTCGGCGAGCATTCGGCCAAAGCCCTAGCCACCGAAGTCGAGAAGATT 251
Db 4475 AGTTCATGGCGCCGCGCTCTTGCTCTGAGCCGACGCTTGGCCACCCCGAGAACAGGT 4416
Qy 252 CCGTAAGAACGCTTCGGCGAGTGGTTG 280
Db 4415 TGGGAAGCTGCGAAGTCCGACGAGATTG 4387

RESULT 13
AAC39454
ID AAC39454 standard; DNA; 1073 BP.
XX
AC AAC39454;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 24688.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic pathway;
KW promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SRP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
```

```
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154773P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157863P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161922P.
PR 28-OCT-1999; 99US-0161933P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 3.5%; Score 33.6; DB 3; Length 1073;
Best Local Similarity 50.0%; Pred. No. 5.8;
Matches 84; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

PR 396 AGCTTCGAGGAATCTCTTCTAGTCTTTCTTGTGATGAAGATCGAAGCGGTTAAAGCGG 455
Db 320 AGTCTTCAATGCTGCTCTTCTTAGCTTTATAGCGGATTCATTTATGAGCTATATGCTCG 379
```

---

```
QY 456 AGCGAAGCTCTCCAGGTCGCGATGATCTCTGTGATCTCCTCGTGGCGAGGAATTTCT 515
Db 380 TAGGCAATTTCTTTCTCCCTCGCTTAGTATTGAAGATTATAACGATCGTGTAGAGCTGT 439
QY 516 GCTCTCGCTGCCGGGAGAGATCCATTGCTGAAGAGATGCTCAAGA 563
Db 440 GGTGCGTGTGAAGCTCAGTATGCTTTGCTGAAGAACTTTGTTGATGA 487

RESULT 14
ABL24038/c
ID ABL24038 standard; DNA; 3780 BP.
XX ABL24038;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 23587.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Claim 1; SEQ ID NO 23587; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 3780 BP; 1059 A; 951 C; 871 G; 899 T; 0 U; 0 Other;

Query Match 3.4%; Score 33.2; DB 4; Length 3780;
Best Local Similarity 45.9%; Pred. No. 15;
Matches 113; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

QY 425 TTGATGAAGCATCGAAGCGGTTAAAGCGCGAGCGAAGCTCTCCAGGTCGCGATGATCT 484
Db 1741 TTCAGGGAGTCCAGCTCGTTGGCATAGTAGCTTCAATGCTTCCAGCCGCGATCCACC 1682
QY 485 CTGTGATCTCCCTGGTGGCGAGGAATTTCTGCTCGCTGCCGGGAGATCCATTGC 544
Db 1681 AGCCGGGAGGACAGGGGTGTCGATGATCTTCTGTCGGCTGGATGCTCACTCCCATGG 1622
QY 545 TGAAGAGATGTCGAAGAGTTGTGCGCGAGTTTGCACCTCGCTCGAGGTCTCTTACC 604
```

Db 1621 TTCAATAGACTGGCTATCTGCTTCTGCTTCCGCTTCACTCGGATTCACAGAGCAAGAGC 1562  
QY 605 TCGGGATACCTCGTGAAGCATTCCTATTTCGAACGAGAGATCTTTGAAGAGGTGCTCG 664  
Db 1561 TCGGAGGTCAAGAGTTCATCTGATCAGATTTCCTCTTCTGTAACGATAGTGTGCTGG 1502  
QY 665 GCCTGA 670  
Db 1501 ACCTGA 1496

RESULT 15  
ABL81387/c  
ID ABL81387 standard; cDNA; 348 BP.  
XX  
AC ABL81387;  
XX  
DT 17-MAY-2002 (first entry)  
XX  
DE Human ovarian cancer related cDNA clone SEQ ID NO:4365.  
XX  
KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200192581-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 29-MAY-2001; 2001WO-US017756.  
XX  
PR 26-MAY-2000; 2000US-0207484P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Algate PA, Harlocker SL, Jones R;  
XX  
DR WPI; 2002-122075/16.  
XX  
Composition for therapy and diagnosis of ovarian cancer comprising  
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
PT polypeptide, antibody specific to polypeptide or T cell expressing  
PT polypeptide.  
XX

Claim 1; SEQ ID NO 4365; 489pp; English.  
XX  
The present invention describes a composition (I) comprising: carriers  
and immunostimulants; and a polypeptide (II) of a ovarian tumour  
polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)  
from the 10912 nucleotide sequences as given in ABL77023 to ABL87934.  
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),  
or antigen presenting cells that express (II). (I) has cytostatic  
activity. An oligonucleotide (IV) that hybridises to (S1) can be used for  
detecting ovarian cancer in a patient's biological sample preferably  
serum or ovarian tissue. The method comprises contacting a biological  
sample from a patient with (IV), detecting the amount of polynucleotide  
hybridising to (IV) and comparing the amount to a predetermined cutoff  
value and thereby detecting ovarian cancer in the patient, where the  
amount of polynucleotide hybridising to (IV) is detected preferably by  
polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is  
useful for stimulating and/or expanding T cells specific for an ovarian  
tumour protein comprising contacting T cells with (III) or (II). (III) is  
useful in design and preparation of ribozyme molecules for inhibiting  
expression of the tumour polypeptides and proteins in tumour cells; and  
to isolate a full length gene from a suitable library e.g., a tumour cDNA  
library using well known techniques  
XX

Sequence 348 BP; 82 A; 102 C; 94 G; 70 T; 0 U; 0 Other;  
Query Match 3.4%; Score 33; DB 6; Length 348;  
Best Local Similarity 55.8%; Pred. No. 5;  
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 105 CGCTGTGACATTGCTCGCGCTTCCGGTTTGAAGCCGGAGATCGCTGGTTGACTCGAC 164  
Db 196 CGCATGGAGGTGCAAGGTGTTCTTCGGGACAGAGCGCAAAATCTGTGTGTGGCTCAAT 137  
QY 165 CACTCCACGCCTTGGCGTTTCAGAAAGATCATGGACTGGTCCGGCGAGCAATTGGG 217  
Db 136 GCCCTCAACCCCTCAGCTCTCAAAAGATCTGGTCTGCCCTCTCTCTCATTTGGG 84

Search completed: October 2, 2004, 00:28:19  
Job time : 419.286 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 1, 2004, 23:40:18 ; Search time 77.5043 Seconds  
(without alignments)  
6959.771 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtgaacagcagtgacggcat.....ttgggcgcgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
C 1	34.8	3.6	1590	US-08-908-758-4	Sequence 4, Appli
C 2	34.8	3.6	1590	US-09-023-587A-4	Sequence 4, Appli
C 3	33.8	3.5	1227	US-09-489-039A-2533	Sequence 2533, Ap
C 4	33	3.4	1102	US-09-620-312D-828	Sequence 828, Ap
C 5	32.8	3.4	1020	US-09-252-991A-9936	Sequence 9936, Ap
C 6	32.8	3.4	1728	US-09-252-991A-9757	Sequence 9757, Ap
C 7	32.8	3.4	1824	US-09-252-991A-10076	Sequence 10076, A
C 8	32	3.3	505	US-09-621-976-15639	Sequence 15639, A
C 9	32	3.3	2898	US-09-489-039A-1679	Sequence 1679, Ap
C 10	32	3.3	5633	US-09-023-655-1490	Sequence 1490, Ap
C 11	31.8	3.3	36941	US-08-311-731A-130	Sequence 130, Ap
C 12	31.6	3.3	3291	US-08-506-296B-3	Sequence 3, Appli
C 13	31.2	3.2	1680	US-09-375-975-39	Sequence 39, Appli
C 14	31.2	3.2	15872	US-09-105-537-1	Sequence 1, Appli
C 15	31.2	3.2	15872	US-09-091-609-1	Sequence 1, Appli
C 16	31.2	3.2	15872	US-09-091-609-3	Sequence 3, Appli
C 17	31.2	3.2	77536	US-09-410-551B-1	Sequence 2, Appli
C 18	31	3.2	2187	US-09-127-219B-2	Sequence 2, Appli
C 19	30.8	3.2	1584	US-09-328-352-775	Sequence 775, Ap
C 20	30.6	3.1	1095	US-09-891-641-46	Sequence 46, Appl
C 21	30.6	3.1	2496	US-09-252-991A-8330	Sequence 8330, Ap
C 22	30.6	3.1	89047	US-09-596-002-34	Sequence 34, Appli
C 23	30.6	3.1	4403765	US-09-103-840A-2	Sequence 2, Appli
C 24	30.6	3.1	4411529	US-09-103-840A-1	Sequence 1, Appli
C 25	30.4	3.1	1802	US-09-221-017B-1019	Sequence 1019, Ap
C 26	30.4	3.1	9686	US-09-221-017B-1003	Sequence 1003, Ap
C 27	30.2	3.1	406	US-08-318-193-52	Sequence 52, Appli

C 28	30.2	3.1	424	2	US-08-765-783A-28	Sequence 28, Appl
C 29	30.2	3.1	424	3	US-08-921-100-28	Sequence 28, Appl
C 30	30.2	3.1	424	3	US-08-880-142-28	Sequence 28, Appl
C 31	30.2	3.1	424	3	US-08-902-201-28	Sequence 28, Appl
C 32	30.2	3.1	424	3	US-09-416-557-28	Sequence 28, Appl
C 33	30.2	3.1	1782	4	US-09-252-991A-15714	Sequence 15714, A
C 34	30.2	3.1	2208	4	US-09-252-991A-15601	Sequence 15601, A
C 35	30.2	3.1	3132	4	US-09-252-991A-11872	Sequence 11872, A
C 36	30	3.1	1002	4	US-09-724-623-28	Sequence 28, Appl
C 37	30	3.1	1023	4	US-09-252-991A-11508	Sequence 11508, A
C 38	30	3.1	1464	4	US-09-252-991A-11733	Sequence 11733, A
C 39	30	3.1	1629	4	US-09-620-312D-715	Sequence 715, App
C 40	30	3.1	1981	4	US-09-132-118-1	Sequence 1, Appli
C 41	30	3.1	2016	3	US-08-444-005-16	Sequence 16, Appli
C 42	30	3.1	2137	1	US-09-161-443-1	Sequence 914, App
C 43	30	3.1	2617	3	US-09-023-655-914	Sequence 2, Appli
C 44	30	3.1	2617	4	US-09-133-944-2	
C 45	30	3.1	9687	3		

#### ALIGNMENTS

##### RESULT 1

US-08-908-758-4/c

; Sequence 4, Application US/08908758

; Patent No. 6429356

; GENERAL INFORMATION:

; APPLICANT: Calgene, Inc.

; TITLE OF INVENTION: METHODS FOR PRODUCING CAROTENOID COMPOUNDS AND SPECIALTY OIL

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Calgene, Inc.

; STREET: 1920 Fifth Street

; CITY: Davis

; STATE: CA

; COUNTRY: USA

; ZIP: 95616

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB

; COMPUTER: IBM PC

; OPERATING SYSTEM: Windows NT 4.0

; SOFTWARE: Microsoft Word For Windows 7.0a

; CURRENT APPLICATION NUMBER: US/08/908,758

; FILING DATE: 8-8-97

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/024,145

; FILING DATE: 8-9-96

; ATTORNEY/AGENT INFORMATION:

; NAME: Donna E. Scherer

; REGISTRATION NUMBER: 34,719

; REFERENCE/DOCKET NUMBER: 36,924

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (916) 753-6313

; TELEFAX: (916) 753-6313

; INFORMATION FOR SEQ ID NO: 4 :

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1590 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

US-08-908-758-4

Query Match 3.6%; Score 34.8; DB 4; Length 1590;  
Best Local Similarity 52.0%; Pred. No. 0.41;  
Matches 102; Conservative 92; Indels 2; Gaps 1;

QY 112 ACATGCTCGCGCTTGGCTTTGAAGCCGGGAGATCGCTGGTGACTCGACCACTCA 171

```
Db 1180 ACGAGTTGGTTGACACAGTTTCAGACCATGAGCTGCTAGACTGCTCGACACCACT 1121
Qy 172 CG--CCTTGGCGTTTCAGAAAGATCATGACTGGTCCGGCGAGCATTTGGGCCAAGCGGTACG 229
Db 1120 GGTCCGGCGCGTCTGCTACATCAACGACGCTCCGAAAGACCTGTCCCGCGCTTACG 1061
Qy 230 CCACCGGAAGTCGAGAAATTCGTAAGAAGACGCTTCGCGAGTGGGTTGATATGCT 289
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCAAGATGCTTCAGAAGTGCATCCACCAAGGTG 1001
Qy 290 TCGCGGTACTTAATGC 305
Db 1000 TTAGATTCATCAGGC 985

RESULT 2
US-09-023-587A-4/c
; Sequence 4, Application US/09023587A
; Patent No. 6653530
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, Christine K.
; APPLICANT: Bhat, Ganesh B.
; APPLICANT: Venkatramesh, Mylavara
; APPLICANT: Rangwala, Shaikat H.
; APPLICANT: Kishore, Ganesh M.
; APPLICANT: Boddupalli, Sekhar S.
; TITLE OF INVENTION: Methods for Producing Carotenoid Compounds, Tocopherol Compounds,
; TITLE OF INVENTION: Specialty Oils in Plant Seeds
; FILE REFERENCE: 16516.122
; CURRENT APPLICATION NUMBER: US/09/023,587A
; CURRENT FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1590
; TYPE: DNA
; ORGANISM: Brassica napus
US-09-023-587A-4

Query Match 3.6%; Score 34.8; DB 4; Length 1590;
Best Local Similarity 52.0%; Pred. No. 0.41;
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

Qy 112 ACATTGCTCGCGTTGCCGGTTTGAAGCGGGAGATCGCTGGTGGTTCAGTCCGACCACTCA 171
Db 1180 ACGAGTTTGGGTTGACGAGTTCGAAGCCATGACTTGTCTAGACTGCTCGACCACT 1121
Qy 172 CG--CCTTGGCGTTTCAGAAAGATCATGACTGGTCCGGCGAGCATTTGGGCCAAGCGGTACG 229
Db 1120 GGTCCGGCGCGTCTGCTACATCAACGACGCTCCGAAAGACCTGTCCCGCGCTTACG 1061
Qy 230 CCACCGGAAGTCGAGAAATTCGTAAGAAGACGCTTCGCGAGTGGGTTGATATGCT 289
Db 1060 GGAGAGTGAACCGTAACAGCTCAATCAAGATGCTTCAGAAGTGCATCCACCAAGGTG 1001
Qy 290 TCGCGGTACTTAATGC 305
Db 1000 TTAGATTCATCAGGC 985

RESULT 3
US-09-489-039A-2533
; Sequence 2533, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Berton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
```

```
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2533
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2533

Query Match 3.5%; Score 33.8; DB 4; Length 1227;
Best Local Similarity 50.2%; Pred. No. 0.75;
Matches 135; Conservative 0; Mismatches 132; Indels 2; Gaps 2;

Qy 1 GTGAACAGCAGTCAGCGCATCGACGGAACGGTAGCAGCATCGATACTGCGGGCGGCTG 60
Db 961 GTGAGTCCATTGATTTAACTCAACGAGGCTTCGCGCTCAGCGCTGCGGCTG 1020
Qy 61 CTAAAGCGTTTGGTTTACGCGCAACGATATACGTCCTGCTAGCGCTGTGACATTGCTC 120
Db 1021 ATGAACGATCTGCGTCTCGACCCGCGCGGTAACTCAACGAGGCTTCGCGCTCAGCGCTGCGGCTG 1080
Qy 121 GCGCTTGGCGGTTTGAAGCCGGGAGATCGCTGGGTTGACTCGACCACTCCACGCTTGGC 180
Db 1081 G-GCCACCGGTTAGCGCGCAGCGGTGCGATTCTGCGGTCAAACTGATGATGAATGGC 1139
Qy 181 GTTCAGAAAGATCATGACTGGTCCGGCGAGCATTTGGGCCAAGCCGTACGCCACCGGAAGT 240
Db 1140 GCGCAGCGCGCAGCAGCACTGGTCTGGTCACCA-TGTGCATCGCGCGCGCCAGGGGATTT 1198
Qy 241 CGAGAAGATTTCGTAAAGAAGCGTTTCG 269
Db 1199 CTATGCTGTTGAGCGGCGAAGCGGCTTAG 1227

RESULT 4
US-09-620-312D-828/c
; Sequence 828, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aiding J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 828
; LENGTH: 1102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (234)..(875)
```

US-09-620-312D-828

Query Match 3.4%; Score 33; DB 4; Length 1102;  
Best Local Similarity 55.8%; Pred. No. 1.3;  
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 105 CGCTGTGACATGTCGCGTTGCGGTTTGAAGCCGGAGATCGCTGGTGTGACTGCAC 164  
D 225 CGCATGAGGTGCAGGTGTCTTCGGGACAGAGCGCAATCTGTGTGCTCAAT 166  
QY 165 CACTCCAGCCTTGGCGTTTCAGAGATCATGACTGTTCGGCGAGATTGGG 217  
D 165 GCCCTCAACCTCAGCTCTCAAAAGATCTGTCTGCCCTCTCTCATTTGGG 113

RESULT 5

US-09-252-991A-9936  
; Sequence 9936, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9936  
; LENGTH: 1020  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9936

Query Match 3.4%; Score 32.8; DB 4; Length 1020;  
Best Local Similarity 45.4%; Pred. No. 1.4;  
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 95 ACCTCGTAGCGTGTGACATTTGCTCGGTTTGAAGCCGGAGATCGCTGGG 154  
D 318 AGGTCCGCTATACGGCGAACCCGACTACTACGCGGCAAGCCGCGATCGACAACCTGG 377  
QY 155 TTGACTCGACCATCTCCACGCTTGGCGTTTCAGAGATCATGACTGGTCCGGCGAGCAT 214  
D 378 TGTTCGCCATTACCTCGATCCCAACGTCCGATGCGATGCGAAGGTCCTGCGCGGAGTGCC 437  
QY 215 GGGCCAAAGCCGTACGCCACCGGAAGTCGAGAAGATTTCCGTAAGAAGACGCTTCGGCAGT 274  
D 438 AGGTCTCGCTGTACCCGAAACCGGAGACGTGCGCGCCTGAAGCAGGACCCGAACTGG 497  
QY 275 GGGTTGATATGCTTCGCGTACTTAATGCGGACAAATTTAAACATCGCCACGAACTCGC 334  
D 498 CGGTGGACGAGATCGATGCCCTGCTGACCACTACATTCGCATCAACACCCAGCAAGC 557  
QY 335 AGCTCAACGAGTACTGCTTG 354  
D 558 CGCTCGACGACCCGCGCGTG 577

RESULT 6

US-09-252-991A-9757  
; Sequence 9757, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9757  
; LENGTH: 1728  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9757

Query Match 3.4%; Score 32.8; DB 4; Length 1728;  
Best Local Similarity 45.4%; Pred. No. 2;  
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 95 ACCTCGTAGCGTGTGACATTTGCTCGGTTTGAAGCCGGAGATCGCTGGG 154  
D 785 AGGTCCGCTATACGGCGAACCCGACTACTACGCGGCAAGCCGCGATCGACAACCTGG 844  
QY 155 TTGACTCGACCATCTCCACGCTTGGCGTTTCAGAGATCATGACTGGTCCGGCGAGCAT 214  
D 845 TGTTCGCCATTACCTCGATCCCAACGTGCGCATGCGAAGGTCCTGCGCGGAGTGCC 904  
QY 215 GGGCCAAAGCCGTACGCCACCGGAAGTCGAGAAGATTTCCGTAAGAAGACGCTTCGGCAGT 274  
D 905 AGGTCTCGCTGTACCCGAAACCGGAGACGTGCGCGCCTGAAGCAGGACCCGAACTGG 964  
QY 275 GGGTTGATATGCTTCGCGTACTTAATGCGGACAAATTTAAACATCGCCACGAACTCGC 334  
D 965 CGGTGGACGAGATCGATGCCCTGCTGACCACTACATTCGCATCAACACCCAGCAAGC 1024  
QY 335 AGCTCAACGAGTACTGCTTG 354  
D 1025 CGCTCGACGACCCGCGCGTG 1044

RESULT 7

US-09-252-991A-10076/c  
; Sequence 10076, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 10076  
; LENGTH: 1824  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10076

Query Match 3.4%; Score 32.8; DB 4; Length 1824;  
Best Local Similarity 45.4%; Pred. No. 2;  
Matches 118; Conservative 0; Mismatches 142; Indels 0; Gaps 0;  
QY 95 ACCTCGTAGCGTGTGACATTTGCTCGGTTTGAAGCCGGAGATCGCTGGG 154  
D 1214 AGGTCCGCTATACGGCGAACCCGACTACTACGCGGCAAGCCGCGATCGACAACCTGG 1155  
QY 155 TTGACTCGACCATCTCCACGCTTGGCGTTTCAGAGATCATGACTGGTCCGGCGAGCAT 214  
D 1154 TGTTCGCCATTACCTCGATCCCAACGTGCGCATGCGAAGGTCCTGCGCGGAGTGCC 1095  
QY 215 GGGCCAAAGCCGTACGCCACCGGAAGTCGAGAAGATTTCCGTAAGAAGACGCTTCGGCAGT 274  
D 1094 AGGTCTCGCTGTACCCGAAACCGGAGACGTGCGCGCCTGAAGCAGGACCCGAACTGG 1035

QY 275 GGGTTGATATGGCTTCGCGTACTTAAATGCGGACAAATTAACATCGCCACGAACTGCG 334  
 Db 1034 CGGTGGACGAGATCGATGCGCTTGTGACCACTTACATTCGCAATCAACACCCAGCAGC 975

QY 335 AGCTCAACGAGTACTGCTTG 354  
 Db 974 CGCTGACGACCGCGCTG 955

RESULT 8  
 US-09-621-976-15639  
 ; Sequence 15639, Application US/09621976  
 ; Patent No. 6639063  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, J.B.  
 ; APPLICANT: Jobert, S.  
 ; APPLICANT: Giordano, J.Y.  
 ; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
 ; FILE REFERENCE: GENSET.054PR2  
 ; CURRENT APPLICATION NUMBER: US/09/621,976  
 ; CURRENT FILING DATE: 2000-07-21  
 ; NUMBER OF SEQ ID NOS: 19335  
 ; SOFTWARE: Patent.pm  
 ; SEQ ID NO 15639  
 ; LENGTH: 505  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-621-976-15639

Query Match 3.3%; Score 32; DB 4; Length 505;  
 Best Local Similarity 10.0%; Pred. No. 1.8;  
 Matches 23; Conservative 111; Mismatches 96; Indels 0; Gaps 0;  
 QY 39 CATCGATATCGCGGCGCTGCTAAAGCGTTTGGTTGACGCGCAAGATATAACGT 98  
 Db 2 MAYMSRYRTSSKRYGKYSYRSMYKYYAWGRKYGTSGRCGSGRMCSKMGY 61  
 QY 99 CGGTAGCGTGTGACATCTCGCGTTCCCGTTTGAAGCGGAGATCGCTGGTTGA 158  
 Db 62 RYSYWGKYSKRYGKYSYRSMYKYYAWGRKYGTSGRCGSGRMCSKMGY 121  
 QY 159 CTCGACCATCCACGCTTGGGCTTCAGAGATCATGACTGGTCCGGCGAGCATTTGGC 218  
 Db 122 KMYYMKYCSASYSYRCKYGTGTRGVGCKRCKSSRYTRYKWTGAC 181  
 QY 219 CAAGCGGTAGCCACCGAGTGCAGAGATTTCCGTAAAGACGCTTC 268  
 Db 182 YGSKGMSCKGSRKYSYRSMYKYYAWGRKYGTSGRCGSGRMCSKMGY 231

RESULT 9  
 US-09-489-039A-1679  
 ; Sequence 1679, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; FILE REFERENCE: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 1679  
 ; LENGTH: 2898  
 ; TYPE: DNA  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-1679

Query Match 3.3%; Score 32; DB 4; Length 2898;  
 Best Local Similarity 48.4%; Pred. No. 5;

Matches 89; Conservative 0; Mismatches 95; Indels 0; Gaps 0;  
 QY 707 TCCATGACGAAGTTTCGTGGTGGCTTTTCCCTTATGAGAGCGCTGAAAAGTAAAGTCGGT 766  
 Db 1337 TCTATGACGTTTCACTGTCGGGTTCTGGCGCAGTATGAAGCGGTGCTGAATAAAAAACGG 1396  
 QY 767 TTGATGAGGAGCGGATCGCAGCCTGCAAGAGCTATTGTTACACCTTCAGCGGTCTAA 826  
 Db 1397 TCTGGGAAGAGCGCAAGAGGAGCATGGGGGAAGAGATTGAAACCCCTCTTCGATCGGCTGA 1456  
 QY 827 TTTTGTAACTGCTTTGAAATCGTGAGTCGATCGTCAGTGGCTCCCTCAGCTGGCTT 886  
 Db 1457 TCTTTGATATCATGACTCCAGTAGCTCGCAGATGCTGCTGATCGCTCATCAGGTT 1516  
 QY 887 GGGG 890  
 Db 1517 GGGG 1520

RESULT 10  
 US-09-023-655-1490  
 ; Sequence 1490, Application US/09023655  
 ; Patent No. 6607879  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cocks, Benjamin G.  
 ; APPLICANT: Susan G. Stuart  
 ; APPLICANT: Jeffrey J. Seilhamer  
 ; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
 ; TITLE OF INVENTION: EXPRESSION  
 ; NUMBER OF SEQUENCES: 1508  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 ; STREET: 3174 PORTER DRIVE  
 ; CITY: PALO ALTO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/023,655  
 ; FILING DATE: HEREWITH  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER:  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Zeller, Karen J.  
 ; REGISTRATION NUMBER: 37,071  
 ; REFERENCE/DOCKET NUMBER: PA-0001 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (650) 855-0555  
 ; TELEFAX: (650) 845-4166  
 ; INFORMATION FOR SEQ ID NO: 1490:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 5633 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: GENBANK  
 ; CLONE: 9862374  
 ; US-09-023-655-1490

Query Match 3.3%; Score 32; DB 4; Length 5633;  
 Best Local Similarity 52.2%; Pred. No. 7.3;  
 Matches 71; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
 QY 823 CTATATTTTGTAAAGCTTTGAAATCGTGAGTCGATGCGTCAAGTGGCTCCCTGAGCTG 882

Db 3 CGAGTGTGGTTCACCTGTGGAGACAGCGGTGGCGAGTGGGTCTCCAGGGCTCTGGGCTG 62  
QY 883 GCTTGGGAACTGAGCGTGGTAGCGGAAGATCCAGACCATCTGATTCACCTTAAACGG 942  
Db 63 GCAAGGCCCGGAGGGGTGGGCGCGAGGAGGCTACAGATCCGCTTCCGCGGCGGG 122  
QY 943 TCTAGATTCTTTGGGC 958  
Db 123 GCGGGTCTTGGGAC 138

## RESULT 11

US-08-311-731A-130  
; Sequence 130, Application US/08311731A  
; Patent No. 6583266  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, DOUGLAS  
; APPLICANT: MAO, JEN-I  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES  
; TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 411  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
; STREET: 600 ATLANTIC AVENUE  
; CITY: BOSTON  
; STATE: MASSACHUSETTS  
; COUNTRY: USA  
; ZIP: 02210  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/311,731A  
; FILING DATE: 530  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: GATES, EDWARD R.  
; REGISTRATION NUMBER: 31,616  
; REFERENCE/DOCKET NUMBER: C0044/7125  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/720-3500  
; TELEFAX: 617/720-2441  
; INFORMATION FOR SEQ ID NO: 130:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 36941 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: circular  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: MYCOBACTERIUM LEPRAE  
US-08-311-731A-130

Query Match 3.3%; Score 31.8; DB 4; Length 36941;  
Best Local Similarity 54.8%; Pred. No. 26;  
Matches 63; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 454 CGAGCGGAAGCTCTCCAGGCTGGATCATCTCTGTCATCTCCCTGGTGGCGAGGAATTT 513  
Db 19602 CGGGAAGTAGTCTCCAGGCTGGACCAATTGATCGAACTCCCGGTGCGGAGACCAA 19661  
QY 514 CTGCTCTCGCTCGCGGAGAAATCCATTCTCGAAGAGATGTCGAAGAGTTTG 568  
Db 19662 CGCACCCGACACGCGCTTTTGGCGGCGCAACAGTCTGATGACCGCTCG 19716

## RESULT 12

US-08-506-296B-3/c  
; Sequence 3, Application US/08506296B  
; Patent No. 6313265  
; GENERAL INFORMATION:  
; APPLICANT: Phillips, Greg  
; APPLICANT: Cunningham, Bruce A.  
; APPLICANT: Crossin, Kathryn L.  
; TITLE OF INVENTION: NEURITE OUTGROWTH-PROMOTING POLYPEPTIDES  
; TITLE OF INVENTION: CONTAINING FIBRONECTIN TYPE III REPEATS AND METHODS OF USE  
; NUMBER OF SEQUENCES: 77  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: The Scripps Research Institute  
; STREET: 10550 No. 6313265th Torrey Pines Road, TPC-8  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: U.S.  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/506,296B  
; FILING DATE: 24-JUL-1995  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fitting, Thomas  
; REGISTRATION NUMBER: 34,163  
; REFERENCE/DOCKET NUMBER: TSRI 488.0  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 554-2937  
; TELEFAX: (619) 554-6312  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3991 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 59..3859  
US-08-506-296B-3  
Query Match 3.3%; Score 31.6; DB 4; Length 3991;  
Best Local Similarity 47.9%; Pred. No. 8.1;  
Matches 91; Conservative 0; Mismatches 99; Indels 0; Gaps 0;  
QY 391 ACGGAAGGCTTCGAGAAATCTCTTTGTAGTCTTTCTTTGATGAAGCATCGAAGCGGTTAA 450  
Db 742 ACTGGGGGCCACGCGGAGGTCGAGGGGCTCCTTCTGATGATGTCGCGGGGCGGAGAA 683  
QY 451 GCGGAGCGGAAGCTCTCCAGGCTGCGATGATCTCTGTGATCTCCTGTGCGGAGAA 510  
Db 682 GTGAGCGTGGCAGATGATGTCGCGGGTGGCTGTCGCCACCATGGCGTTGAGAAGTAGAG 623  
QY 511 TTTCTCTCTCGCTCGCGGAGAAATCCATTCTCGAAGAGATGTCGAAGAGTTTGTG 570  
Db 622 GTTCCCATCTGCCCCATAGAGACCCCTCTCGTCTGAGCATGTGAACGATTCGCTGTT 563  
QY 571 CGCGGATTTG 580  
Db 562 GAGCCAATAG 553  
RESULT 13  
US-09-375-975-39/c  
; Sequence 39, Application US/09375975  
; Patent No. 6593116  
; GENERAL INFORMATION:

; APPLICANT: Huismann, Gjalte W.  
; APPLICANT: Peoples, Oliver P.  
; APPLICANT: Skraly, Frank A.  
; TITLE OF INVENTION: Transgenic Microbial Polyhydroxyalkanoate Producers  
; FILE REFERENCE: MBX 033  
; CURRENT APPLICATION NUMBER: US/09/375,975  
; CURRENT FILING DATE: 1999-08-17  
; EARLIER APPLICATION NUMBER: 60/096,852  
; EARLIER FILING DATE: 1998-09-18  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 39  
; LENGTH: 1680  
; TYPE: DNA  
; ORGANISM: Pseudomonas oleovorans  
US-09-375-975-39

Query Match 3.2%; Score 31.2; DB 4; Length 1680;  
Best Local Similarity 44.2%; Pred. No. 6.6;  
Matches 129; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 181 GTTCAGAGATCATGACTGTCGGCGAGCATNTGGCCCAAGCCGTACGCCACCGGAAGT 240  
Db |||||  
504 GTTGACCAAGTCTTTGGCAGGTTGGACAGGCCATCGACAGGCTCTTGCCCGCGGTTTC 445  
QY 241 CGAAGAGATTTCCGTAAGAAGACGCTTCGGCAGTGGTTGATAATGGCTTCGCCGTAATT 300  
Db |||||  
444 GAAGAAGCGCTTGAATGCTGCGGGTTGGACAGAGGTTTGGTCGAGGCATGGCTTCGGT 385  
QY 301 AATCGGACAAATTTAAACATCGCCAGCAATCGCAGCTCAACGAGTACTGCTTGTCTGAC 360  
Db |||||  
384 CATCAGTTGATCAGCACTGCGCGGCTGATGTCCTGGGCGGACAGTCTGTTGCC 325  
QY 361 GAAGCATTACAGCGCTAAGGCATATGAACGGAAGGCTTCGAGGAATCTTGTGTAGTC 420  
Db |||||  
324 GATCCAGTCTCAGCTCTTGGCCAGCCAGATAGTTTGCAGGTAGCGCGGTAAAG 265  
QY 421 TTCTTGTGATGAGCTCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGG 472  
Db |||||  
264 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGACGCTGCTGCTTTCCGGG 213

RESULT 14  
US-09-105-537-1/c  
; Sequence 1, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105,537A  
; CURRENT FILING DATE: 1998-06-26  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-1

Query Match 3.2%; Score 31.2; DB 3; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25;  
Matches 129; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 181 GTTCAGAGATCATGACTGTCGGCGAGCATTTGGCCCAAGCCGTACGCCACCGGAAGT 240  
Db |||||  
14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGCCGCGGTTTC 14592  
QY 241 CGAAGAGATTTCCGTAAGAAGACGCTTCGGCAGTGGTTGATAATGGCTTCGCCGTAATT 300  
Db |||||

Query Match 3.2%; Score 31.2; DB 3; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25;  
Matches 129; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 181 GTTCAGAGATCATGACTGTCGGCGAGCATTTGGCCCAAGCCGTACGCCACCGGAAGT 240  
Db |||||  
14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGCCGCGGTTTC 14592  
QY 241 CGAAGAGATTTCCGTAAGAAGACGCTTCGGCAGTGGTTGATAATGGCTTCGCCGTAATT 300  
Db |||||

Db 14591 GAAGAAGCGTTTGAATGCTGCGCGGTTGGACAGAGGTTGTGTCGAGCAATGGCTTCGGT 14532  
QY 301 AATCGGACAAATTTAAACATCGCCAGCAATCGCAGCTCAACGAGTACTGCTTGTCTGAC 360  
Db |||||  
14531 CATCAGTTGATGAGCAACTGCGCGGCTGATGTCCTGGGCGGACAGGTGCTGTTGCC 14472  
QY 361 GAAGCATTACAGCGCTTAAGGGCATATGAACGGAAGGCTTCGAGGAATCTTGTGTAGTC 420  
Db |||||  
14471 GATCCAGTCTCAGCTCTTTCGGCCAGCCAGATAGGTTTGCAGGTAGCGCGGTAAAG 14412  
QY 421 TTCTTGTGATGAAGCATCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGG 472  
Db |||||  
14411 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGACGCTGCTGCTTCCGGG 14360

RESULT 15  
US-09-091-609-1/c  
; Sequence 1, Application US/09091609  
; Patent No. 6600029  
; GENERAL INFORMATION:  
; APPLICANT: SHERMAN, DAVID H.  
; APPLICANT: WILLIAMS, MARK D.  
; APPLICANT: XUE, YONGQUAN  
; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
; TITLE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES  
; FILE REFERENCE: 600.297US2  
; CURRENT APPLICATION NUMBER: US/09/091,609  
; CURRENT FILING DATE: 1998-06-19  
; EARLIER APPLICATION NUMBER: PCT/US96/20119  
; EARLIER FILING DATE: 1996-12-18  
; EARLIER APPLICATION NUMBER: 60/008,847  
; EARLIER FILING DATE: 1995-12-19  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 15872  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (20)....(13909)  
US-09-091-609-1

Query Match 3.2%; Score 31.2; DB 4; Length 15872;  
Best Local Similarity 44.2%; Pred. No. 25;  
Matches 129; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 181 GTTCAGAGATCATGACTGTCGGCGAGCATTTGGCCCAAGCCGTACGCCACCGGAAGT 240  
Db |||||  
14651 GTTGACCAAGTCTTTGGCCAGGTTGGACAGGCCATCGACAGGCTCTTGCCGCGGTTTC 14592  
QY 241 CGAAGAGATTTCCGTAAGAAGACGCTTCGGCAGTGGTTGATAATGGCTTCGCCGTAATT 300  
Db |||||  
14591 GAAGAAGCGTTTGAATGCTGCGCGGTTGACAGGGTTGTCGAGAGCCATGGCTTCGGT 14532  
QY 301 AATCGGACAAATTTAAACATCGCCAGCAATCGCAGCTCAACGAGTACTGCTTGTGTAGTC 360  
Db |||||  
14531 CATCAGTTGATGACGAACCTGCGCGGCTGATGTCCTGGGCGGACAGGTGCTGTTGCC 14472  
QY 361 GAAGCATTACAGCGCTTAAGGCATATGAACGGAAGGCTTCGAGGAATCTTGTGTAGTC 420  
Db |||||  
14471 GATCCAGTCTCAGCTCTTTCGGCCAGCCAGATAGGTTTGCAGGTAGCGCGGTAAAG 14412  
QY 421 TTCTTGTGATGAAGCATCGAAGCGGTTAAAGCGGAGCGGAGCTCTCCAGG 472  
Db |||||  
14411 TGGGTTGTTGCTCCATGCGCGGTTCATTGAAGCGACGCTGCTGCTTCCGGG 14360

Search completed: October 2, 2004, 05:16:08  
Job time : 83.5043 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 2, 2004, 03:13:49 ; Search time 504.178 Seconds  
(without alignments)  
9773.579 Million cell updates/sec

Title: US-10-668-047-3

Perfect score: 972

Sequence: 1 gtaacagcagtgacgcat.....ttggccgtacgaacgttag 972

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 3340653 seqs, 2534783454 residues

Total number of hits satisfying chosen parameters: 6681306

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*

2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*

3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*

4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*

6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*

7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*

8: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*

9: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*

10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*

11: /cgn2\_6/ptodata/2/pubpna/US09D\_PUBCOMB.seq:\*

12: /cgn2\_6/ptodata/2/pubpna/US09E\_PUBCOMB.seq:\*

13: /cgn2\_6/ptodata/2/pubpna/US09F\_PUBCOMB.seq:\*

14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*

15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*

16: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*

17: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq:\*

18: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq:\*

19: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	38.6	4.0	1163020	16	US-10-398-221-10
C 2	38.6	4.0	3011208	16	US-10-398-221-10
C 3	38.2	3.9	521	13	US-10-424-599-91751
C 4	37.8	3.9	802	15	US-10-184-644-312
C 5	37.8	3.9	802	15	US-10-184-644-312
C 6	36	3.7	2061	16	US-10-369-493-33642
C 7	35.8	3.7	78953	16	US-10-085-117-31
C 8	35.2	3.6	653	15	US-10-184-644-402
C 9	35.2	3.6	653	15	US-10-184-644-402
C 10	34.8	3.6	1590	14	US-10-041-472-4
C 11	34.4	3.5	1376	15	US-10-241-596-71
C 12	34.4	3.5	1376	15	US-10-241-596-73
C 13	34.4	3.5	1376	15	US-10-241-596-75
C 14	34.4	3.5	1376	15	US-10-241-596-77

C 15	34.4	3.5	1385	15	US-10-241-596-69	Sequence 69, Appl
C 16	34.4	3.5	2600	15	US-10-241-596-33	Sequence 33, Appl
C 17	34.4	3.5	2600	15	US-10-241-596-35	Sequence 35, Appl
C 18	34.4	3.5	2600	15	US-10-241-596-37	Sequence 37, Appl
C 19	34.4	3.5	2609	15	US-10-241-596-31	Sequence 31, Appl
C 20	34.4	3.5	2619	15	US-10-241-596-152	Sequence 152, App
C 21	34.4	3.5	2625	15	US-10-241-596-150	Sequence 150, App
C 22	34.4	3.5	2625	15	US-10-241-596-164	Sequence 164, App
C 23	34.4	3.5	2625	15	US-10-241-596-166	Sequence 166, App
C 24	34.4	3.5	2637	15	US-10-241-596-168	Sequence 168, App
C 25	34.4	3.5	2637	15	US-10-241-596-156	Sequence 156, App
C 26	34.4	3.5	2640	15	US-10-241-596-29	Sequence 29, Appl
C 27	34.4	3.5	2643	15	US-10-241-596-61	Sequence 61, Appl
C 28	34.4	3.5	2643	15	US-10-241-596-158	Sequence 158, App
C 29	34.4	3.5	3382	15	US-10-241-596-160	Sequence 160, App
C 30	34.4	3.5	3382	15	US-10-241-596-39	Sequence 39, Appl
C 31	34.4	3.5	3382	15	US-10-241-596-49	Sequence 49, Appl
C 32	34.4	3.5	3382	15	US-10-241-596-53	Sequence 53, Appl
C 33	34.4	3.5	3382	15	US-10-241-596-57	Sequence 57, Appl
C 34	34.4	3.5	3388	15	US-10-241-596-41	Sequence 41, Appl
C 35	34.4	3.5	3388	15	US-10-241-596-47	Sequence 47, Appl
C 36	34.4	3.5	3388	15	US-10-241-596-51	Sequence 51, Appl
C 37	34.4	3.5	3391	15	US-10-241-596-55	Sequence 55, Appl
C 38	34.4	3.5	3391	15	US-10-241-596-43	Sequence 43, Appl
C 39	34.4	3.5	3397	15	US-10-241-596-45	Sequence 45, Appl
C 40	34.2	3.5	448	13	US-10-424-599-77596	Sequence 77596, A
C 41	34.2	3.5	777	15	US-10-184-644-348	Sequence 348, App
C 42	34.2	3.5	777	15	US-10-184-644-348	Sequence 348, App
C 43	34	3.5	653	17	US-10-767-701-10779	Sequence 10779, A
C 44	33.8	3.5	4298	13	US-10-424-599-70962	Sequence 70962, A
C 45	33.8	3.5	7133	17	US-10-432-241-1	Sequence 1, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-398-221-10/c  
; Sequence 10, Application US/10398221  
; Publication No. US20040018514A1  
; GENERAL INFORMATION:  
; APPLICANT: KUNST, Frederik  
; APPLICANT: GLASER, Philippe  
; TITLE OF INVENTION: Listeria innocua, genome and applications  
; FILE REFERENCE: 344 702 - US  
; CURRENT APPLICATION NUMBER: US/10/398,221  
; CURRENT FILING DATE: 2003-03-27  
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061  
; PRIOR FILING DATE: 2001-10-04  
; PRIOR APPLICATION NUMBER: FR 00/12 697  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 4025  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 10  
; LENGTH: 1163020  
; TYPE: DNA  
; ORGANISM: Listeria innocua  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(end)  
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u  
US-10-398-221-10

Query Match 4.0%; Score 38.6; DB 16; Length 1163020;  
Best Local Similarity 61.4%; Pred. No. 2;  
Matches 62; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY	453	GGAGCGGAGCTCCAGGTCGATGATCTCTGCGATCTCCCTGTCGCGAGGAAAT	512
Db	513671	GCAGCGCAAAATTTGGCAGCGTCGATGAAATTTATCAAGTTGGCGGTGCGCAGCAATC	513612
QY	513	TCTGCTCGCTCGCGGCGGCGAATCCATTCGTGAAGAGA	553



## RESULT 6

```
QY 716 AAGTTCGGTGGCTTTCTTATGAGCGCGTGAAGAGTAAGGTCGCTTTCATGAGG 775
Db 10022 ATGTAATTCACAGCTTGTGCTGTGGGAAGTCCAGCAAGAAAGGTCACCTTTATGGG 9963
QY 776 AGCGGCATCGC 786
Db 9962 AGCAAGAC 9952

RESULT 8
US-10-184-644-402
; Sequence 402, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-402

Query Match 3.6%; Score 35.2; DB 15; Length 653;
Best Local Similarity 6.4%; Pred. No. 0.7;
Matches 28; Conservative 141; Mismatches 266; Indels 0; Gaps 0;

QY 11 GTGACGGCATCGACGAGCGTAGCGACATCGATCTACTGCGGGCGCTGCTAAAGCGTT 70
Db 88 SEINEDSDHYAIMPPLQFMEIPSDMRRELFFRDIERGDIVIGRISSIREFGFFMWLIC 147
QY 71 TTGGGTTTACGCGCAACGATATAACGTCGCTAGCGCTGTGACATTCGTCGCGCTTGCGG 130
Db 148 LGSGIMRDIHLEITALEPLRDVPSHNSHGDPLSYQTGDIIRAGIKIDIRYHEKLAIVSL 207
QY 131 GTTGAACGCGGAGATCGCTGGTGTGACTCGACCATCCAGCCTCCGCTTGCAGAGA 190
Db 208 YSSLPPLHLSGILKGVISSEELPLYRRSVLSNSLESYENVMQSSILGFVNPVVEFLL 267
QY 191 TCATGACTGTGCGCGAGCATTCGGCCAGCGGTACGCCACCGAGTCGAGAAGATT 250
Db 268 EKLGDENPPLMRLGQSKNFSEDDFASALRKKQSASWALKCVKIGVDYFKVGRHVDAM 327
QY 251 TCCGTAAGAAGACGCTTCGGCAGTGGTGTGATAATGGCTTCGCCGTACTTAATGCGACA 310
Db 328 NEYNKALEIDKQNVREALVARGALYATKGSNLKAIEDFELALENCPTHNRNARYLCQTLVE 387
QY 311 ATTTAAACATCCCGCAGAACTCGCAGCTCAACGAGTACTGCTGTGTGACGAAGCATTAC 370
Db 388 RGGQLEEBEKFNLAEYSYKKALALDETFFKDAEDALQKLHKVMQKSLREKQAEKEKQK 447
QY 371 AGCGCGTAAGGCATATGAACGAGGCTTCAGAAATCTCTGTAGCTTTCTTGTATG 430
Db 448 TKIETSAREKRLKLEKRLKKRKRKSTSSSVSSADESVSSSSSSSGHRRHKHKKR 507
QY 431 AAGCATCGAAGCGG 445
Db 9962 AGCAAGAC 9952

RESULT 10
US-10-041-472-4/c
; Sequence 4, Application US/10041472
; Publication No. US20020092039A1
```

```
Db 508 NRSESSSRRRHSSR 522

RESULT 9
US-10-184-634-402
; Sequence 402, Application US/10184634
; Publication No. US200300684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 402
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-402

Query Match 3.6%; Score 35.2; DB 15; Length 653;
Best Local Similarity 6.4%; Pred. No. 0.7;
Matches 28; Conservative 141; Mismatches 266; Indels 0; Gaps 0;

QY 11 GTGACGGCATCGACGAGCGTAGCGACATCGATCTACTGCGGGCGCTGCTAAAGCGTT 70
Db 88 SEINEDSDHYAIMPPLQFMEIPSDMRRELFFRDIERGDIVIGRISSIREFGFFMWLIC 147
QY 71 TTGGGTTTACGCGCAACGATATAACGTCGCTAGCGCTGTGACATTCGTCGCGCTTGCGG 130
Db 148 LGSGIMRDIHLEITALEPLRDVPSHNSHGDPLSYQTGDIIRAGIKIDIRYHEKLAIVSL 207
QY 131 GTTGAACGCGGAGATCGCTGGTGTGACTCGACCATCCAGCCTCCGCTTGCAGAGA 190
Db 208 YSSLPPLHLSGILKGVISSEELPLYRRSVLSNSLESYENVMQSSILGFVNPVVEFLL 267
QY 191 TCATGACTGTGCGCGAGCATTCGGCCAGCGGTACGCCACCGAGTCGAGAAGATT 250
Db 268 EKLGDENPPLMRLGQSKNFSEDDFASALRKKQSASWALKCVKIGVDYFKVGRHVDAM 327
QY 251 TCCGTAAGAAGACGCTTCGGCAGTGGTGTGATAATGGCTTCGCCGTACTTAATGCGACA 310
Db 328 NEYNKALEIDKQNVREALVARGALYATKGSNLKAIEDFELALENCPTHNRNARYLCQTLVE 387
QY 311 ATTTAAACATCCCGCAGAACTCGCAGCTCAACGAGTACTGCTGTGTGACGAAGCATTAC 370
Db 388 RGGQLEEBEKFNLAEYSYKKALALDETFFKDAEDALQKLHKVMQKSLREKQAEKEKQK 447
QY 371 AGCGCGTAAGGCATATGAACGAGGCTTCAGAAATCTCTGTAGCTTTCTTGTATG 430
Db 448 TKIETSAREKRLKLEKRLKKRKRKSTSSSVSSADESVSSSSSSSGHRRHKHKKR 507
QY 431 AAGCATCGAAGCGG 445
Db 508 NRSESSSRRRHSSR 522

RESULT 10
US-10-041-472-4/c
; Sequence 4, Application US/10041472
; Publication No. US20020092039A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Shewmaker, Christine
; TITLE OF INVENTION: METHODS FOR PRODUCING CAROTENOID COMPOUNDS AND SPECIALITY OILS IN SEEDS
; TITLE OF INVENTION: SEEDS
; FILE REFERENCE: 16516.141
; CURRENT APPLICATION NUMBER: US/10/041.472
; CURRENT FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 08/908758
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/024145
; PRIOR FILING DATE: 1996-09-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1590
; TYPE: DNA
; ORGANISM: Brassica napus
US-10-041-472-4

Query Match      3.6%; Score 34.8; DB 14; Length 1590;
Best Local Similarity 52.0%; Pred. No. 1.5;
Matches 102; Conservative 0; Mismatches 92; Indels 2; Gaps 1;

QY 112 ACATTGCTCGCGTTGCGGTTTGAACCGGGAGATCGCTGGGTGACTCGACCACTCCA 171
DB 1180 ACGGAGTTGGGTTGACGAGTTCGAAGCCATGGACTGTCTAGACTGCTCGACACCACT 1121

QY 172 CG--CCTTGGCGTTTCAGAACATCATGACTGGTCCGCGGAGCATTTGGCCCAAGCCGTACG 229
DB 1120 GGTCCGCGCGCTGCTGTCTACATCAACACGCGTCCGAAAGGACCTGTCCCGGCTTACG 1061

QY 230 CCACCGGAAGTCGAGAGATTTCCGTAAGAGACGCTTCGCGCAGTGGGTGTGATAATGGCT 289
DB 1060 GGAGAGTGAACCGTAAACAGCTCAATCCAGATGCTTCAGAGTGCATCACCAACGGTG 1001

QY 290 TCGCCGCTACTTAATGC 305
DB 1000 TTAGATTCCATCAGGC 985

RESULT 11
US-10-241-596-71/c
; Sequence 71, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-71

Query Match      3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAAAGTTCTGGTGGTCTTTCCTTATGGAGCGCGTGAAGAAGT 756
DB 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGCGCGTTAAAT 1204

QY 757 AAAGGTCCGTTTGTATGAGGAGCGCATCGACGCTTCAAGAGCTATTTCGTTACACCTTCA 816
DB 1203 AGCAGCCAGGTTGGTGTACGAGGTTGAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

QY 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
DB 1143 CACGATGTTGATTTTGAATACGCTTG 1116

RESULT 13
US-10-241-596-75/c
; Sequence 75, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
```

```
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAAAGTTCTGGTGGTCTTTCCTTATGGAGCGCGTGAAGAAGT 756
DB 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGCGCGTTAAAT 1204

QY 757 AAAGGTCCGTTTGTATGAGGAGCGCATCGACGCTTCAAGAGCTATTTCGTTACACCTTCA 816
DB 1203 AGCAGCCAGGTTGGTGTACGAGGTTGAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

QY 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
DB 1143 CACGATGTTGATTTTGAATACGCTTG 1116
```

```
RESULT 12
US-10-241-596-73/c
; Sequence 73, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 73
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-73
```

```
Query Match      3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 697 GACCTTATTCTCCATGACGAAAGTTCTGGTGGTCTTTCCTTATGGAGCGCGTGAAGAAGT 756
DB 1263 GAAGTTTTTTCAGTTTTTGTGAAGTTCATGTTGTGATTTCCGTGTTCTGCGCGTTAAAT 1204

QY 757 AAAGGTCCGTTTGTATGAGGAGCGCATCGACGCTTCAAGAGCTATTTCGTTACACCTTCA 816
DB 1203 AGCAGCCAGGTTGGTGTACGAGGTTGAACCATCGTAGATAGTGTAGTTAACTTCGG 1144

QY 817 GCGGCTCTAAATTTTGTAAACTGCTTTG 844
DB 1143 CACGATGTTGATTTTGAATACGCTTG 1116
```

```
RESULT 13
US-10-241-596-75/c
; Sequence 75, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
```

```
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-75

Query Match          3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 697 GACCTTATTCCTCCATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGCGTGAAGT 756
Db 1263 GAAGTTTTTTCAGTTTGTGAAGTTCATGTTGTTGTTTCCGTTCTGCGCGTTAAAT 1204

Qy 757 AAAGTCCGTTTGTATGAGGCGGCATCGCAGCTTTCCTTATGAGGCGCGTGAAGT 756
Db 1203 AGCAGCAGGTTGGTTTACGAGTTGAACCATCGTAGATGTTAACTTCGG 1144

Qy 817 GCGGCTCTAAATTTTGTAACTGCTTTG 844
Db 1143 CACGATGTTGATTTTGAATCTGCTTG 1116

RESULT 14
US-10-241-596-77/c
; Sequence 77, Application US/10241596
; Publication No. US20030166238A1
; GENERAL INFORMATION:
; APPLICANT: Microbiological Research Authority
; APPLICANT: The Speywood Laboratory Limited
; TITLE OF INVENTION: Recombinant Toxin Fragments
; FILE REFERENCE: 1581.0130003
; CURRENT APPLICATION NUMBER: US/10/241,596
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 09/255,829
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: US 09/242,689
; PRIOR FILING DATE: 1999-02-23
; PRIOR APPLICATION NUMBER: PCT/GB97/02273
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 08/782,893
; PRIOR FILING DATE: 1996-12-27
; PRIOR APPLICATION NUMBER: GB 9625996.5
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: GB 9617671.4
; PRIOR FILING DATE: 1996-08-23
; NUMBER OF SEQ ID NOS: 175
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Clostridium botulinum
US-10-241-596-69

Query Match          3.5%; Score 34.4; DB 15; Length 1385;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 697 GACCTTATTCCTCCATGACGAGTTCGTGGTGGCTTTTCTTATGAGGCGCGTGAAGT 756
Db 1263 GAAGTTTTTTCAGTTTGTGAAGTTCATGTTGTTGTTTCCGTTCTGCGCGTTAAAT 1204

Qy 757 AAAGTCCGTTTGTATGAGGCGGCATCGCAGCTTTCCTTATGAGGCGCGTGAAGT 756
Db 1203 AGCAGCAGGTTGGTTTACGAGTTGAACCATCGTAGATGTTAACTTCGG 1144

Qy 817 GCGGCTCTAAATTTTGTAACTGCTTTG 844
Db 1143 CACGATGTTGATTTTGAATCTGCTTG 1116
```

Search completed: October 2, 2004, 08:33:28  
Job time : 510.178 secs

```
US-10-241-596-77
Query Match          3.5%; Score 34.4; DB 15; Length 1376;
Best Local Similarity 52.0%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 11.1792 Seconds  
(without alignments)  
2779.248 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIDGTVASIDTARAL.....EDPDHLHLNGSRFLGPIYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	632.5	37.4	546	2 AF2268	type II site-speci
2	603.5	35.7	316	2 S35516	type II site-speci
3	511	30.3	326	1 NDOFS	type II site-speci
4	104	6.2	772	1 JQ2024	outer layer protei
5	100	5.9	772	1 JQ2025	outer layer protei
6	100	5.9	772	2 A46110	outer capsid prote
7	99.5	5.9	1203	2 H87687	helicase, UvrD/Rep
8	97.5	5.8	289	2 B83189	elongation factor
9	97	5.7	485	2 E83663	glutamyl-tRNA synt
10	96	5.7	516	2 H72427	DNA mismatch repai
11	95.5	5.7	797	2 G83570	probable two-compo
12	95.5	5.7	1057	2 F89892	carbamoyl-phosphat
13	95	5.6	409	2 AD3118	ABC transporter, s
14	95	5.6	409	2 C98169	hypothetical prote
15	95	5.6	4574	2 G02520	plectin - human
16	95	5.6	4684	2 A59404	plectin [imported]
17	94.5	5.6	574	2 T08434	misato protein - f
18	94.5	5.6	619	2 A69352	conserved hypothet
19	94	5.6	264	2 AD3353	probable membrane
20	94	5.6	312	2 S76937	hypothetical prote
21	94	5.6	772	1 A44052	outer layer protei
22	93.5	5.5	359	2 A96038	probable transcrip
23	93.5	5.5	399	2 B69069	GTP-binding protei
24	93	5.5	772	2 A46108	outer capsid prote
25	92.5	5.5	449	2 D75033	probable prolifera
26	92	5.4	3951	1 VFHBI	Fl protein - avian
27	91.5	5.4	318	2 F84495	probable Athila re
28	91.5	5.4	788	1 JVDLHH	DNA-directed DNA p
29	91.5	5.4	1360	2 JCS839	GTP-N protein - h

30	90.5	5.4	406	2 H70364	iron-sulfur cofact
31	90.5	5.4	456	2 F47677	phosphomannomutase
32	90	5.3	474	2 T19219	hypothetical prote
33	90	5.3	480	1 NIAIMA	nitrogenase (EC 1.
34	90	5.3	497	2 AF1986	nitrogenase molybd
35	90	5.3	852	2 T28790	hypothetical prote
36	89.5	5.3	360	2 A75315	probable chalcone
37	89.5	5.3	1010	2 T36383	probable large ATP
38	89	5.3	666	2 S75289	ribonuclease II -
39	88.5	5.2	489	1 NIBCAT	nitrogenase (EC 1.
40	88.5	5.2	504	2 S05197	signal recognition
41	88.5	5.2	504	2 S54143	SRP 54 protein - h
42	88	5.2	367	2 E83676	pyruvate dehydroge
43	88	5.2	408	2 H90222	GTP binding protei
44	88	5.2	772	1 JQ2026	outer layer protei
45	88	5.2	1965	2 S75200	fat protein - Syne

ALIGNMENTS

RESULT 1

AF2268

type II site-specific deoxyribonuclease [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AF2268

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anab

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AF2268

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-546 <R>

A;Cross-references: GB:BA000019; PIDN:BA075400.1; PID:gl7132835; GSPDB:GNO0179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr3701

Query Match 37.4%; Score 632.5; DB 2; Length 546;

Best Local Similarity 40.2%; Pred. No. 1.5e-42;

Matches 129; Conservative 65; Mismatches 122; Indels 5; Gaps 3;

QY	4	SDGIDGTVAS--IDTARALLKRFGFDQRYNVRSAVTLLALAGLKPGRWVDSTTPRLGV	61
DB	223	SEAIKAVLAAKRIEALAILKDISAPKEQNSALFLLALADIRPEIPWTQATSPRRRI	282
QY	62	OKIMDSGEHWAQPYATGSRDPKKTILQWVDNGFAVLNAMNLTATNSQLNEYCLSD	121
DB	283	TEMMDWFRDHYGQYAPNTRVRRQTMQFVQMGIWENPDQDFRINSPKWCYQLHQ	342
QY	122	ALQALRAYTEGFEGLSVFLDEASKAVKARAEALQAAISVDLPGGEFFLSPAGONPL	181
DB	343	FVTLKSYSGEWEETRRNYISVKMLQDRNRI--PMIPVSLPNGQAIQSSGGNQL	400
QY	182	LKKMVEEFVPRFAPRSTVLYLGDTRGKSLFEREIFEVEVLGLTFDPRGMPDILHDEVR	241
DB	401	IKEILLENFCPRFTPEGVLVFDGAGNKFINVETQKFEI-GIELDPHGKMDIVVYERQ	459
QY	242	GHLFLMEAVKSGPDDEHRSLQELFVTPSAGLIPVNCFNENRSMQWLPELAMETEA	301
DB	460	EWLVLIEAVTSTGPNVNLKRNELKRLFQSSRQGLFVFTAFPSRKEMTRYLABISWET	519
QY	302	VAEDPDHLHLNGSRFLGPIYE	322
DB	520	VAAQPDHMLHNGERFLGPIYE	540

RESULT 2

S35516



outer capsid protein VP4 - human rotavirus (strain 116E)  
C:Species: human rotavirus  
C:Date: 07-Apr-1994 #sequence\_revision 18-Nov-1994 #text\_change 20-Sep-1999  
C:Accession: A46110  
R:Gentsch, J.R.; Das, B.K.; Jiang, B.; Bhan, M.K.; Glass, R.I.  
Virology 194, 424-430, 1993  
A:Title: Similarity of the VP4 protein of human rotavirus strain 116E to that of the bovine  
A:Reference number: A46110; MUID:93242780; PMID:8386888  
A:Contents: 116E  
A:Accession: A46110

QY 160 MISVDLPGGEFLLSPAGON-----PLKKMVEEFVPRFAPRSTVLYLGD-----RGKH 209



QY 275 LIFVNC-----FENRESM-----RQWLPELAW-----TEA 300  
Db 287 VLEIENPEELDNHPSKIWKFLNEEKVKKSLEETLKRNLARKWYSVAYEISSRAL 346  
QY 301 WVAEDDPH 308  
Db 347 SVAEAPSH 354

RESULT 11  
G83570  
Probable two-component sensor PA0600 [imported] - Pseudomonas aeruginosa (strain PA01)  
C;Species: Pseudomonas aeruginosa  
C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C;Accession: G83570  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: G83570  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-797 <STO>  
A;Cross-references: GB:AE004496; GB:AE004091; NID:g9946468; PIDN:AAG03989.1; GSPDB:GN001187  
A;Experimental source: strain PA01  
C;Genetics:  
A;Gene: PA0600

Query Match 5.7%; Score 95.5; DB 2; Length 797;  
Best Local Similarity 24.7%; Pred. No. 17;  
Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;

QY 10 TVASIDTARALKRFGDAQ-RYNVRSAYTLA-----LAGLKPG---DRWVDST 55  
Db 325 TQFSDNSTLGLIUVNWDSEVRVYANAAEQMLGHADQQLVDFLADFEPLDMRWLNL- 383  
QY 56 TPRGLGVQKIMDSGEHWAKPYATGSRSE---DFRKKTLRQWVDNGFAVLNADLNINATNSQ 112  
Db 384 -----WRR---ARNSEGLSPFTRCLR-----AD----- 405  
QY 113 LNEYCLSDALQALRAYGTGEPESLVVFLDEASKAVKARAEALQ---AAM---ISVDLPGL 167  
Db 406 -GSWLPADVLSFLR-FGT---SEYLWVFLSDVTERRAR-EALQSEARMKGIASNVPG 459  
QY 168 GEEFLSPAGQNPLLKKMYEEFV-PPFAPRSTVLYLGDTRGKHSLP---EREIFEVLGL 223  
Db 460 -----MVFLRRPRAGAFSDPAYISE--GSEALVGYSAARELIESGRGI 500  
QY 224 TEDPHGRMPDLILHDEVRCWLFMEAVKSGKPGFDEHRHRSRLQELFVTPSAGLIFVN--- 279  
Db 501 RGLVH---PD---DEERYWSSQMAI-----DENRDHWQGRILTRQELRWADIKAS 547  
QY 280 --CFENRESMROMLPELAWETEA 301  
Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 12  
F89892  
carbamoyl-phosphate synthase large chain [imported] - Staphylococcus aureus (strain N315)  
C;Species: Staphylococcus aureus  
C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C;Accession: F89892  
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogudma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A;Reference number: A89758; MUID:21311952; PMID:11418146  
A;Accession: F89892  
A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-1057 <KUR>  
A;Cross-references: GB:BA000018; PID:g13701002; PIDN:BAB42298.1; GSPDB:GN00149  
A;Experimental source: strain N315  
C;Genetics:  
A;Gene: pyrAB  
C;Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain; biotin c

Query Match 5.7%; Score 95.5; DB 2; Length 1057;  
Best Local Similarity 21.5%; Pred. No. 26;  
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;

QY 82 EDRFKTKLRQWVDNGFAVLNADLN-----IATNSQLN--EY-CLSDALQALR 127  
Db 214 KEIEYEVWRKDNDAIVVCMNENIDPVGHITGDSIVVAPSQTLSDVEYQMLRDVSLKVIR 273  
QY 128 AYCTEG-----FRESLVVFLDEASKAVKARAEALQ-----AAMISVDLPGL 168  
Db 274 ALGIEGGCNVQLALDPHSFDYIIIEVNPV-SSSSALASKATGYPIAKLAIAVGLTLD 332  
QY 169 EEFLLSP-----AGQNPLKKWVEFVPRF-----APRSTVLYLGDTR 206  
Db 333 E--MLNPITGTSYAAPEFTLDYVISK-IPRFPDFKFEKGERELGTQMKATGEYMAIGRT- 388  
QY 207 GKHSLPEREIFEVLGLTLD-----PHGRMPDL-----ILHDEVRCWLFMEAVKSKG 254  
Db 389 -----YEESLLKAIRSLEVGVHGLPNGESFLLDYIKERISHQDDERLFFIGEAIR--- 440  
QY 255 PFDEERHRSRLQELFVTPSAGLIFVNCFENRESMROMLPE 293  
Db 441 ----RGTTLEETHNMTQIDYFELHKKFQNIIDIEHQKE 474

## RESULT 13

AD3118  
ABC transporter, substrate binding protein Atu4568 [imported] - Agrobacterium tumefaciens  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C;Accession: AD3118  
R;Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; erage, G.; Gallet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, F star, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AD3118  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-409 <KUR>  
A;Cross-references: GB:AE008689; PIDN:AAL45362.1; PID:g17743057; GSPDB:GN00187  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: Atu4568  
A;Map position: linear chromosome

Query Match 5.6%; Score 95; DB 2; Length 409;  
Best Local Similarity 26.0%; Pred. No. 7.4;  
Matches 56; Conservative 23; Mismatches 62; Indels 74; Gaps 12;

QY 51 WVDSTTPRLGVQ-KIMDSGEHW-----AKPYATGSRDEPRK-----KTLRQ----- 91  
Db 193 WAENKPKRLGVTLPAKGGSGGGFIYSVALNYLTG---DCRKQLTDYNTQLOQAEDWAMES 249  
QY 92 -----W-----VDNGFAVLNADLNINATNSQLNEVCLSDALQALRAYCTEGF 134  
Db 250 ECLTPWDDYRLLLAELTNG-----NADTLNINQQLYMTGVTWED---QVMSFLGTROL 303  
QY 135 EBSLVVFLDEASKA-----VKARAEALQAMISVDLPFGGEEFLL-----SPACQ 178  
Db 304 PDSFRVTLLEKQGVGSGDAMFVPANAKNVASALLIDMAMSKFQWLKLENKASRPRD 363

```
QY 179 --NPLKKWVEEFV-----PRFAPRSTVL 200
Db 364 VTNDLMPKAVQEHVLPQSVYPRLSLPFAFWDMSAL 398

RESULT 14
C98169
hypothetical protein AGR_L_608 [imported] - Agrobacterium tumefaciens (strain C58, Cerec
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: C98169
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: C98169
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-409 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK88877.1; PID:g15158644; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_608
A;Map position: linear chromosome

Query Match 5.6%; Score 95; DB 2; Length 409;
Best Local Similarity 26.0%; Pred. No. 7.4;
Matches 56; Conservative 23; Mismatches 62; Indels 74; Gaps 12;

QY 51 WVDSTTPRLGVQ-KIMDWSGEHW-----AKPYATGSRDEFK-----KTLRQ----- 91
Db 193 WAENKPKELGVTLPKAGSGGGFIYSVALNLTG---DCRKQITDYNQTLQQAEDWAMES 249

QY 92 -----W-----VDNGFAVLNADNINIATNSQINCYCLSDALQALRAYGTGTF 134
Db 250 ECLTPWMDYRRLLAVBELTNG---NADTLNINQQLYMGTVWED--QVMSFLGTQKL 303

QY 135 EESLVVFLDEASKA-----VKARAEALQAMISVDLPGGEEFLL-----SPAGQ 178
Db 304 PDSFRVTLLEKGVGSGDAMFVPANAKNVASALLIDMANSKEFQWKLKNSRSPRTD 363

QY 179 --NPLKKWVEEFV-----PRFAPRSTVL 200
Db 364 VTNDLMPKAVQEHVLPQSVYPRLSLPFAFWDMSAL 398

RESULT 15
G02520
plectin - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 20-Aug-1999
C;Accession: G02520
R;McLean, W.H.I.; Smith, F.J.D.
submitted to the EMBL Data Library, March 1996
A;Reference number: H01385
A;Accession: G02520
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: mRNA
A;Residues: 1-4574 <MCL>
A;Cross-references: EMBL:U53204; NID:g1477645; PIDN:AA05427.1; PID:g1477646
C;Genetics:
A;Gene: PLECI
C;Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein 8
P;68-283/Domain: alpha-actinin actin-binding domain homology <ACT>

Query Match 5.6%; Score 95; DB 2; Length 4574;
Best Local Similarity 25.9%; Pred. No. 2.2e+02;
Matches 62; Conservative 39; Mismatches 70; Indels 68; Gaps 14;

QY 84 FRKKTIRQ--WVDNGFAVLNADNINI-ATNSQINCYCLSDALQALRAYGTGTFEE---- 136
Db 2200 FAEQTLRQKAQVEQLTTL---RLQLEETDHQKN---LLDEELQRLKAEATEAQRQSV 2253
```

```
QY 137 -----SLVVFIDEASKAVKARAEALQAMISVDLPGGEEFLLSPAGQNPLLLKKWVEEFV 191
Db 2254 EEELFSVRVQMEELSK-LKARIEAENRALILRDKDNTQRFLOEEAEK---MKQVAEE--- 2306

QY 192 RFAPRSTVLVLDTRGKHSIFEREIFEVLGLITFDPHGRWEDLLIHDEVRGWLFLMEAVK 251
Db 2307 --AARLSV-----AAQFAARLQLABEDLA-----QQRALAEKMLKEKMQA---VQEA 2351

QY 252 SKGPPD-----EERHRSLOELFVTPSAGLIFVNCFNENRSMRWMLPELAWETEAW 301
Db 2352 LKAEAEILLQOKKELAQEQARRLQ-----EDKEQMAQ---QLAEETQGF 2391

Search completed: October 1, 2004, 16:11:04
Job time : 15.1792 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 7.58591 Seconds  
(without alignments)  
2217.092 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIDGTVASIDTARAL.....EDPDHLHLNGSRFLGPFYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	603.5	35.7	316	1	T2BB_BACSU
2	511	30.3	326	1	T2PS_PROST
3	104	6.2	772	1	VP4_ROTBA
4	101.5	6.0	1057	1	CARB_STAEP
5	101	6.0	772	1	VP4_ROTBB
6	100	6.0	772	1	VP4_ROTBU
7	97.5	5.8	289	1	EFTS_PSEAE
8	97	5.7	485	1	SYE_BACHD
9	96.5	5.7	513	1	C942_VICSA
10	96	5.7	1057	1	MUTL_THEMA
11	95.5	5.7	1057	1	CARB_STAAM
12	95.5	5.7	1057	1	CARB_STAAM
13	95	5.6	4584	1	PLEL_HUMAN
14	94.5	5.6	454	1	TOLB_BRAJA
15	94.5	5.6	619	1	Y817_ARCFU
16	93	5.5	1723	1	A1M1_HUMAN
17	92.5	5.5	492	1	Y193_COREF
18	92	5.4	447	1	DNAE_ENTEA
19	92	5.4	6629	1	R1AB_IBVB
20	92	5.4	6629	1	R1AB_IBVCB
21	91.5	5.4	788	1	DPOL_HPBHE
22	91.5	5.4	1360	1	MSH6_HUMAN
23	90.5	5.4	456	1	MANB_SALMO
24	90	5.3	497	1	N1FD_ANASP
25	89.5	5.3	485	1	RNS_RHILLO
26	89	5.3	666	1	R2H2_SYNY3
27	88.5	5.2	489	1	N1FD_THIFE
28	88.5	5.2	504	1	SR54_HUMAN
29	88	5.2	480	1	SYE_FASMU
30	88	5.2	553	1	5NTD_VIBVU
31	88	5.2	772	1	VP4_ROTBB
32	88	5.2	2483	1	COA2_HUMAN
33	87.5	5.2	504	1	SR54_MOUSE

34 87.5 5.2 701 1 PALL\_ORYSA  
35 87.5 5.2 872 1 SYA\_LACIA  
36 87 5.2 466 1 SYC\_BACHD  
37 87 5.2 737 1 RECG\_MYCTU  
38 87 5.2 1215 1 ATC6\_YEAST  
39 86.5 5.1 1136 1 CED2\_HUMAN  
40 86.5 5.1 2105 1 POLR\_ASGVP  
41 86.5 5.1 2805 1 MAPA\_HUMAN  
42 86 5.1 436 1 TOLB\_RHILLO  
43 86 5.1 447 1 HSLD\_PSEAE  
44 86 5.1 500 1 N1FD\_RHICP  
45 86 5.1 879 1 SYA\_CLOPE

## ALIGNMENTS

RESULT 1  
T2BB\_BACSU  
ID T2BB\_BACSU STANDARD; PRT; 316 AA.  
AC P33562; 1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Type II restriction enzyme BsuBI (EC 3.1.21.4) (Endonuclease BsuBI)  
DE (R.BsuBI)  
GN HSDR OR HSRB.  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=ISB8;  
RX MEDLINE=93126092; PubMed=1480472;  
RA Xu G.-L., Kapfer W., Walter J., Trautner T.A.;  
RT "BsuBI -- an isospecific restriction and modification system of PstI:  
characterization of the BsuBI genes and enzymes."  
RL Nucleic Acids Res. 20:6517-6523(1992).  
CC -!- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE CTGCAG AND  
CLEAVES AFTER A-5'.  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give  
specific double-stranded fragments with terminal 5'-phosphates.  
CC -!- COFACTOR: Magnesium.  
CC -!- SUBUNIT: Homodimer.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
CC EMBL; L01541; AAA18170.1; --  
DR PIR; S35516; S35516.  
DR REBASE; 616; BsuBI.  
KW Restriction system; Hydrolase; Nuclease; Endonuclease; Magnesium.  
SQ SEQUENCE 316 AA; 36197 MW; B1C2E45851E7F3C9 CRC64;

Query Match 35.7%; Score 603.5; DB 1; Length 316;  
Best Local Similarity 37.1%; Pred. No. 6.1e-40;  
Matches 117; Conservative 65; Mismatches 132; Indels 1; Gaps 1;  
QY 8 DTVASIDTARALLKRFQDAQRYNVRSAVTLALAGLKPGRWVDSTTPRLGVQKIMDW 67  
Db 3 EGMHNVKAEIKLGLPKQGNRSALCLSLMNITQDKTWEAESPLGIIPMEF 62  
QY 68 SEHWAKPYATGSRDFRKKTLRQWVDNGFAVLNADNLNATNSQNLNCLSEALQALR 127  
Db 63 CRINYGKAYPNRSRFRFTMHQFVAGIALYNPDKTRPVNSPKAVYQIEATLELTK 122  
QY 128 AYTGFGEESLVVFLDEASKAVKARAEALQAAMISVDLPGGEEFLSPAGQNPLLKQVVE 187

Db 123 CYNTEWSELLARYLSNRQTLIVRYAKEROQNKIPVQIAEGKEIYITPGEHSELKAIIE 182  
Qy 188 EFVPRFAPRSTVLYLGDTRGHSLFEREIEFEVLGLTFDPHGRMPDLILHDEVGWLFLM 247  
Db 183 EFAPRYVPGRLIYAGDTGERXGFDDELLAQ-LGVWIDSHGKMPDVVIYPEKKWLLI 241  
Qy 248 EAVKSGPDEERHRSLOELFVTPSAGLIFVNCENRESMOWLPDLAWETEAWEADPD 307  
Db 242 ESVTSHGPDVHKRHEELAKFNGTAGIVYTAPNKLMLARYLNNISWETEVWVADAPS 301  
Qy 308 HLIHLNGSRFLGPYE 322  
Db 302 HLIHFGVRLGPYE 316

RESULT 2  
ID T2PS\_PROST STANDARD; PRT; 326 AA.  
AC P00640;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Type II restriction enzyme PstI (EC 3.1.21.4) (Endonuclease PstI)  
(R.PstI).  
GN PSTII.  
OS Providencia stuartii.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Providencia.  
OX NCBI\_TaxID=588;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=164;  
RX MEDLINE=84239756; PubMed=6330092;  
RA Walder R.Y., Walder J.A., Donelson J.E.;  
RT "The organization and complete nucleotide sequence of the PstI  
restriction-modification system.";  
RL J. Biol. Chem. 259:8015-8026(1984).  
CC -!- FUNCTION: RECOGNIZES THE DOUBLE-STRANDED SEQUENCE CTCGAG AND  
CLEAVES AFTER A-5'.  
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage of DNA to give  
specific double-stranded fragments with terminal 5'-phosphates.  
CC -!- COFACTOR: Magnesium.  
CC -!- SUBUNIT: Homodimer.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; K02081; AAA25673.1; -.  
DR PIR; A00783; NDOFS.  
DR REBASE; 1536; PstI.  
KW Restriction system; Hydrolase; Nuclease; Endonuclease; Magnesium.  
SQ SEQUENCE 326 AA; 37413 MW; 475841183B414504 CRC64;

Query Match 30.3%; Score 511; DB 1; Length 326;  
Best Local Similarity 35.7%; Pred. No. 1.1e-32;  
Matches 112; Conservative 53; Mismatches 141; Indels 8; Gaps 3;

Qy 17 ABALLKRFQDAQRNVNSAVTLLALAGLKPGRWVDSVTPRLGVQKIMDWSGEHWKPY 76  
Db 9 AKEILKALGPLPQQVNDNRSGWLLALANIKPDSWKAKAPLLPTVSMFEIRTEYKDY 68  
Qy 77 ATGSEDFRKKTLROWDNGFVAINADNLNIATNSQINLEYCLSDALQALRAYTGEFEE 136  
Db 69 KNSSETIRQTLHQEQEQAIVDRNEDLPSRAINSKDNYSLNQVIIDILHNPNGNWE 128  
Qy 137 SLIVFLDEASKAVARAELQAAMISVDLPGGEEFLSPAGQNPLLKQWVEEFVPRF-AP 195  
Db 129 LIQQFLTHVPSLQELIYERALARDRIPKILDLDTQISLSFGEHNLHADIHVFCRPFVD 188

Qy 196 RSTVLYLGDT-----GKHSLEFEREIEFEVLGLTFDPHGRMPDLILHDEVGWLFLM 249  
Db 189 MGKILYIGDTASSNEGGKMLVLDSEYLLK-LGVPPMSHDKLPDVVVYDEKRWLFLEA 247  
Qy 250 VKSGPDEERHRSLOELFVTPSAGLIFVNCENRESMOWLPDLAWETEAWEADPDHL 309  
Db 248 VTHGPISPKRWLEALSSCTVGKVVYTAFTTRTFRKNAANIATWETVMIADNPDM 307  
Qy 310 IHLNGSRFLGPYE 323  
Db 308 VHFNGDRFLGPHDK 321

RESULT 3  
ID VP4\_ROTBA STANDARD; PRT; 772 AA.  
AC P36307;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)  
DE [Contains: Outer capsid proteins VP5 and VP8].  
GN S4.  
OS Bovine rotavirus (serotype 10 / strain A44).  
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.  
OX NCBI\_TaxID=36437;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93286580; PubMed=8398807;  
RA Taniguchi K., Urasawa T., Urasawa S.;  
RT "Independent segregation of the VP4 and the VP7 genes in bovine  
rotaviruses as confirmed by VP4 sequence analysis of G8 and G10  
bovine rotavirus strains.";  
RL J. Gen. Virol. 74:1215-1221(1993).  
CC -!- SUBCELLULAR LOCATION: Outer capsid.  
CC -!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other  
product is VP5.  
CC -!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; D13392; -; NOT ANNOTATED CDS.  
DR InterPro; IPR000416; Cap VP4.  
DR InterPro; IPR008985; ConA\_like\_lec\_g1.  
DR Pfam; PF00426; VP4; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 772 OUTER CAPSID PROTEIN VP4.  
FT CHAIN 1 242 OUTER CAPSID PROTEIN VP8.  
FT CHAIN 249 772 OUTER CAPSID PROTEIN VP5.  
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 772 AA; 86619 MW; D5FE1201624FE994 CRC64;

Query Match 6.2%; Score 104; DB 1; Length 772;  
Best Local Similarity 22.9%; Pred. No. 2;  
Matches 77; Conservative 47; Mismatches 128; Indels 84; Gaps 16;  
Qy 1 VNSSDGDIGCTVASIDTARALLKRFQDAQRNVNSAVTLLALA-----GLKPGDRWVDS 55  
Db 473 VFSNDDYQTFPIANSVTRQDLER-QLDEMRRREFNELSANIALSLQIDLALLP----LDMF 527  
Qy 56 TPRGLGVQKIMDWSGEHWKAPYATGSRDFRK-----KTLRQ 91

```

Db 528 SMFSGIOSTVE-----AAKTFATSVMKFKKSLAKSVNSLTDAITDAAGSISRSSTLRS 582
QY 92 -----WVNGFAVLNADNINATNSQLNECYCLSDALQALRAYGTEGP--EESLVVFL 142
Db 583 VNSVASVWTDLSIDVSTDNVVAAT-----ATAAKKFRVKEFTTEFNGVSFD 630
QY 143 DEASKAVKARAEALQAAMISVDLPFGGEEFLLSPAGQNPLLKKMVEEVPVPFAPR---STV 199
Db 631 DISAAVVKTKMKNLV---VD-----BELL-----PQITEASEKIPNRAVRLIDGK 676
QY 200 LYLGDTRKSHLPEREIFEVEVLGTFDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258
Db 677 VYEVTEGKYFAVLTETFEVW---FDAE-RFAELVTDQSIVSAIDFKTIKNLNDNYGI 732
QY 259 ERHRSQELFVTPSAGLIFVNCENRSMROWLPEL 294
Db 733 TREQALNMLRSDPKVLRSFIN---QNNPIKNRIEQL 766

RESULT 4
CARB_STAEP STANDARD; PRT; 1057 AA.
ID CARB_STAEP STANDARD; PRT; 1057 AA.
AC Q8CPJ4; 2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR SE0879.
OS Staphylococcus epidermidis.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y.-Q., Ren S.-X., Li H.-L., Wang Y.-X., Fu G., Yang J.,
RA Qin Z.-Q., Miao Y.-G., Wang W.-Y., Chen R.-S., Shen Y., Chen Z.,
RA Yuan Z.-H., Zhao G.-P., Ou D., Danchin A., Wen Y.-M.;
RT "Genome-based analysis of virulence genes in a non-biofilm-forming
RT Staphylococcus epidermidis strain (ATCC 12228).";
RL Mol. Microbiol. 49:1577-1593(2003).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
CC -!- PATHWAY: Arginine biosynthesis.
CC -!- SUBUNIT: Pyrimidine biosynthesis; first step.
CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -!- SIMILARITY: Belongs to the carb family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; AE016746; AAC004476.1; -
DR HAMAP; MF_01210; -; 1
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR004362; MGS_like.
DR InterPro; IPR000169; Shprot_acsite.
DR Pfam; PF00289; CPSase_L_chain; 2.
DR Pfam; PF02786; CPSase_L_D2; 2.

```

```

DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF02142; MGS_L_1; -; 1.
DR PRINTS; PRO0098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.
FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.
FT REPEAT 1 546
FT REPEAT 547 1057
FT NP_BIND 153 210 ATP (POTENTIAL).
FT NP_BIND 302 352 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).
FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DB1CAE59 CRC64;

Query Match 6.0%; Score 101.5; DB 1; Length 1057;
Best Local Similarity 22.7%; Pred. No. 4.8;
Matches 61; Conservative 40; Mismatches 83; Indels 85; Gaps 14;

QY 82 EDRFKTKLROWDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIEYEMRDKNDAIVVNCNMENIDPVGIHTGDSIVVAPSQTLSDVEYQMLRDVSLKVR 273
QY 128 AYCTEG-----FEESLVVFLDEASKAVKAAEALQ-----AAMISVDLP 168
Db 274 ALGIEGCVQLALDPSLNYIIIEVNPV--SRSSALASKATGYPIAKLAIAVGLTLD 332
QY 169 EEFLSP-----AGQNPLLKMWEEFVPRF-----APRSTVLYLGDTR 206
Db 333 E--MLNPITGTSVAAFEPTLDYVISK-IPFPDPKPKGERELGTQMKATGEVWAGRT- 388
QY 207 GKHSLEFERIFEVLGLTDPH-----GRMPDL-----ILHDEVRGMLFLMEAVKSKG 254
Db 389 -----YEESLLKAIRSLEYGVHHLGSLNGESYELDVIKERIGHQDDERLFFIGEAIR--- 440
QY 255 PDDEERHRSLOELFVTPSAGLIFVNCFN 283
Db 441 -----RGTSLLELHNMTKIDYFPLNKFQN 464

RESULT 5
VP4_ROTBB
ID VP4_ROTBB STANDARD; PRT; 772 AA.
AC F35746;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)
DE [Contains: Outer capsid proteins VP5 and VP8].
GN S4.
OS Bovine rotavirus (serotype 10 / strain B223).
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10930;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033121; PubMed=1329320;
RA Hardy M.E.; Gorziglia M.; Woode G.N.;
RT "Amino acid sequence analysis of bovine rotavirus B223 reveals a
RT unique outer capsid protein VP4 and confirms a third bovine VP4
RT type.";
RL Virology 191:291-300(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93286580; PubMed=8389807;
RA Taniguchi K., Urasawa T., Urasawa S.;

```

"Independent segregation of the VP4 and the VP7 genes in bovine rotaviruses as confirmed by VP4 sequence analysis of G8 and G10 bovine rotavirus strains.";  
J. Gen. Virol. 74:1215-1221(1993).  
-!- SUBCELLULAR LOCATION: Outer capsid.  
-!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other product is VP5.  
-!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; M92986; AAA02978.1; -;  
DR EMBL; D13394; -; NOT\_ANNOTATED\_CDS.  
DR PIR; A44052; A44052.  
DR PIR; JQ2025; JQ2025.  
DR InterPro; IPR000416; Cap VP4.  
DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
DR Pfam; PF00426; VP4; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 772 OUTER CAPSID PROTEIN VP4.  
FT CHAIN 1 242 OUTER CAPSID PROTEIN VP5.  
FT CHAIN 249 772 OUTER CAPSID PROTEIN VP5.  
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 2 2 R -> A (IN REF. 2).  
FT CONFLICT 193 193 N -> H (IN REF. 2).  
FT CONFLICT 535 535 S -> T (IN REF. 2).  
SQ SEQUENCE 772 AA; 86540 MW; 61FA433F71079A8D CRC64;  
  
Query Match 6.0%; Score 101; DB 1; Length 772;  
Best Local Similarity 22.9%; Pred. No. 3.5;  
Matches 77; Conservative 46; Mismatches 129; Indels 84; Gaps 16;  
  
QY 1 VNSSDGDGTVASIDTARALLKRFQDQRYNVRSAVTLALA-----GLKPGDRWVDST 55  
DB 473 VPSNDYQTPIANSVTVVQDLER-QLDEMRRFELSANIALSOLIDLALLP-----LDMF 527  
QY 56 TPLRGVQKIMWSGHEHWAKPYATGSRDPRK-----KTLRQ 91  
DB 528 SMFSGIQSTVE-----AAKTFATSVMKFRKSLAKSVNSLTDAITDAAGSISRSTLRS 582  
QY 92 -----WVDNGFAVLNADNLNIATNSQLNEYCLSDALQALRAYGTEGF--EESLVVPL 142  
DB 583 VNSAASVMTDISDVIDSTDNVVAAT-----ATAAAKFRVKEFTTEFGVSPD 630  
QY 143 DEAKAVKARABALQAAMISVDLPGEFLLSPAGQNPLLKKWVEEFVPRFAPR---STV 199  
DB 631 DLISAAVTKMSKLVN-----VD-----EELL-----PQITEASEKFIENRAYRLIDGK 676  
QY 200 LYLGTGRGKHSIFEREIFEVVLGLTFDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258  
DB 677 VYEVTEGKYFAYLTETFEVW---FDAB-REAELVTDSPVISAIDFTIKNLNDNYGI 732  
QY 259 EHRHSLOELFVTPSAGLIFVNCFNRESMROWLPEL 294  
DB 733 TREQALNMLRSPKVLRSFIN--QNNPIIKNRIEQL 766  
  
RESULT 6  
VPA\_ROTU  
ID\_VPA\_ROTU STANDARD; PRT; 772 AA.  
AC Q09113;  
DT 01-NOV-1995 (Rel. 32, Created)

01-NOV-1995 (Rel. 32, Last sequence update)  
10-OCT-2003 (Rel. 42, Last annotation update)  
Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)  
[Contains: Outer capsid proteins VP5 and VP8].  
GN S4.  
OS Human rotavirus (serotype G / strain 116E).  
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.  
OX NCBI\_TaxID=42567;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93242780; PubMed=8386888;  
RA Gentsch J.R., Das B.K., Jiang B., Bhan M.K., Glass R.I.;  
RT "Similarity of the VP4 protein of human rotavirus strain 116E to that of the bovine B223 strain.";  
RL Virology 194:424-430(1993).  
CC -!- SUBCELLULAR LOCATION: Outer capsid.  
CC -!- PTM: VP8 is one of two trypsin cleavage products of VP4; the other product is VP5.  
CC -!- SIMILARITY: Belongs to the rotaviruses VP4 protein family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; L07934; AAA47357.1; -;  
DR PIR; A46110; A46110.  
DR InterPro; IPR000416; Cap VP4.  
DR InterPro; IPR008985; ConA\_like\_lec\_gl.  
DR Pfam; PF00426; VP4; 1.  
KW Coat protein; Glycoprotein.  
FT CHAIN 1 772 OUTER CAPSID PROTEIN VP4.  
FT CHAIN 1 242 OUTER CAPSID PROTEIN VP5 (BY SIMILARITY).  
FT CHAIN 249 772 OUTER CAPSID PROTEIN VP5 (BY SIMILARITY).  
FT CARBOHYD 32 32 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 123 123 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 173 173 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 772 AA; 86882 MW; 31139CD20695F822 CRC64;  
  
Query Match 5.9%; Score 100; DB 1; Length 772;  
Best Local Similarity 22.6%; Pred. No. 4.2;  
Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;  
  
QY 1 VNSSDGDGTVASIDTARALLKRFQDQRYNVRSAVTLALA-----GLKPGDRWVDST 55  
DB 473 VPSNDYQTPIANSVTVVQDLER-QLDEMRRFELSANIALSOLIDLALLP-----LDMF 527  
QY 56 TPLRGVQKIMWSGHEHWAKPYATGSRDPRK-----KTLRQ 91  
DB 528 SMFSGIRSTIE-----AAKNFATSVMKFRKSNLAKSVNSLTDAITDAAGSISRSTLRS 582  
QY 92 -----WVDNGFAVLNADNLNIATNSQLNEYCLSDALQALRAYGTEGF--EESLVVPL 142  
DB 583 ANSASVMTDISDVIDSTDNVVAAT-----ATAAAKFRVKEFTTEFGVSPD 630  
QY 143 DEAKAVKARABALQAAMISVDLPGEFLLSPAGQNPLLKKWVEEFVPRFAPR---STV 199  
DB 631 DLISAAVTKMSKLVN-----VD-----EELL-----PQITEASEKFIENRAYRLIDGK 676  
QY 200 LYLGTGRGKHSIFEREIFEVVLGLTFDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258  
DB 677 VYEVTEGKYFAYLTETFEVW---FDAB-REAELVTDSPVISAIDFTIKNLNDNYGI 732  
QY 259 EHRHSLOELFVTPSAGLIFVNCFNRESMROWLPEL 294  
DB 733 TREQALNMLRSPKVLRSFIN--QNNPIIKNRIEQL 766







QY 169 BEFLSP-----AGQPLKKMVEEFVPR-----APRSTVLYLGDT 206  
 Db 333 E--MLNPITGTSYAAFEFLDYVSK-LPRPFDKFKGERELGTQMKATGEVMAIGRT- 388  
 QY 207 GKSLFEREIEFVGLGTFD-----PHGRMPDL-----ILHDEVRWGLFMEAVKSG 254  
 Db 389 -----YBSLLKAIRSLVGVHHLGPNESFDLDYIKERISHQDDERLFFIGEAIR--- 440  
 QY 255 PFDERHRSLOELFVTPSAGLIFVNCFFENRESMROWLPE 293  
 Db 441 -----RGTTLEIHNMTQIDYFFLHKFQNIIDIEHOLKE 474

## RESULT 12

CARB\_STRAW STANDARD; PRT; 1057 AA.

AC P58940;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-phosphate synthetase ammonia chain).  
 DE CARB OR PYRAB OR MW1086.  
 GN Staphylococcus aureus (strain MW2).  
 OS Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OC NCBI\_TaxID=196620;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22040717; PubMed=12044378;  
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A., Nagai Y., Iwano K., Asano K., Naimi T., Kuroda H., Cui L., Yamamoto K., Hiramatsu K.;  
 RA "Genome and virulence determinants of high virulence community-acquired MRSA."  
 RT Lancet 359:1819-1827 (2002).  
 RL -! CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP + phosphate + L-glutamate + carbamoyl phosphate.  
 CC -! COFACTOR: Binds 3 manganese ions per subunit (By similarity).  
 CC -! PATHWAY: Arginine biosynthesis.  
 CC -! PATHWAY: Pyrimidine biosynthesis; first step.  
 CC -! SUBUNIT: Composed of two chains; the small (or glutamine) chain promotes the hydrolysis of glutamine to ammonia, which is used by the large (or ammonia) chain to synthesize carbamoyl phosphate (By similarity).  
 CC -! SIMILARITY: Belongs to the carb family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF004825; BAB94951.1; -.  
 DR HAMAP; MF 01210; -; 1.  
 DR InterPro; IPR006275; CarA\_L\_glu.  
 DR InterPro; IPR005483; CPase\_L.  
 DR InterPro; IPR005479; CPase\_L\_D2.  
 DR InterPro; IPR005480; CPase\_L\_D3.  
 DR InterPro; IPR005481; CPase\_L\_N.  
 DR InterPro; IPR004362; MGS like.  
 DR Pfam; PF02789; CPase\_L\_chain; 2.  
 DR Pfam; PF02786; CPase\_L\_D2; 2.  
 DR Pfam; PF02787; CPase\_L\_D3; 1.  
 DR Pfam; PF02142; MGS; 1.  
 DR PRINTS; PRO0098; CPASE.  
 DR TIGRfams; TIGR01369; CPaseII\_lrg; 1.  
 DR PROSITE; PS00866; CPASE\_1; 2.  
 DR PROSITE; PS00867; CPASE\_2; 2.  
 DR Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;  
 KW ATP-binding; Manganese; Complete proteome.  
 FT DOMAIN 1 401 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.

FT DOMAIN 402 546 OLIGOMERIZATION DOMAIN.  
 FT DOMAIN 547 929 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.  
 FT DOMAIN 930 1057 ALLOSTERIC DOMAIN.  
 FT REPEAT 1 546  
 FT REPEAT 547 1057  
 FT NP\_BIND 153 210 ATP (POTENTIAL).  
 FT NP\_BIND 302 352 ATP (POTENTIAL).  
 FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).  
 FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).  
 FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).  
 FT METAL 820 820 MANGANESE 3 (BY SIMILARITY).  
 FT METAL 832 832 MANGANESE 3 (BY SIMILARITY).  
 SQ SEQUENCE 1057 AA; 117185 MW; D8E3B09F9BC6F152 CRC64;

Query Match 5.7%; Score 95.5; DB 1; Length 1057;

Best Local Similarity 21.5%; Pred. No. 14;  
 Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;  
 QY 82 EDRKKTLRWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
 Db 214 KEIEYEVNRKNDNAIVVCNMENIDPVGIHTGDSIVAPQSOTLSDVEYQMLRDLVKVIR 273  
 QY 128 AVCTEG-----FEESLVVFLDEASKAVKARALQ-----AAMISVDLPGG 168  
 Db 274 ALGIEGGCNVOLALDPHPSFDYIIIEVNPV--SRSSALASKATGYPIAKLAIAVGLTLD 332  
 QY 169 BEFLSP-----AGQPLKKMVEEFVPR-----APRSTVLYLGDT 206  
 Db 333 E--MLNPITGTSYAAFEFLDYVSK-LPRPFDKFKGERELGTQMKATGEVMAIGRT- 388  
 QY 207 GKSLFEREIEFVGLGTFD-----PHGRMPDL-----ILHDEVRWGLFMEAVKSG 254  
 Db 389 -----YBSLLKAIRSLVGVHHLGPNESFDLDYIKERISHQDDERLFFIGEAIR--- 440  
 QY 255 PFDERHRSLOELFVTPSAGLIFVNCFFENRESMROWLPE 293  
 Db 441 -----RGTTLEIHNMTQIDYFFLHKFQNIIDIEHOLKE 474

## RESULT 13

ID PLEI HUMAN STANDARD; PRT; 4684 AA.  
 AC Q15149; Q15149; Q16640;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HDI).  
 GN PLECI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX TISSUE=Placenta;  
 RX MEDLINE=96210632; PubMed=8633055;  
 RA Liu C.-G., Maercker C., Caстанon M.J., Hauptmann R., Wiche G.;  
 RT "Human plectin: organization of the gene, sequence analysis, and chromosome localization (8q24)."  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.  
 RX MEDLINE=96312447; PubMed=8698233;  
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B., Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K., McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M., Uitto J.;  
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy: cDNA cloning and genomic organization."  
 RL Genes Dev. 10:1724-1735 (1996).  
 RN [3]  
 RP VARIANT MD-EBS 1003-GIN--ALA-1005 DEL.  
 RX MEDLINE=97049959; PubMed=8694687;

RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,  
RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.,  
RA "Homozygous deletion mutations in the plectin gene (PLECI) in patients  
RT with epidermolysis bullosa simplex associated with late-onset  
RT muscular dystrophy.";  
RL Hum. Mol. Genet. 5:1539-1546(1996).  
RN [4]  
RP VARIANT MD-EBS LEU-429 INS.  
RX MEDLINE=21090821; PubMed=1115198;  
RA Bauer J.W., Rouan F., Kofler B., Rezniczek G.A., Kornacker I.,  
RA Muss W., Hametner R., Klausberger A., Huber A., Pohl-Gubo G.,  
RA Wiche G., Uitto J., Hintner H.;  
RT "A compound heterozygous one amino-acid insertion/nonsense mutation in  
RT the plectin gene causes epidermolysis bullosa simplex with plectin  
RT deficiency.";  
RL Am. J. Pathol. 158:617-625(2001).  
RN [5]  
RP VARIANT EBS1 TRP-2110.  
RX MEDLINE=21841370; PubMed=11851880;  
RA Koss-Harnes D., Hoeyheim B., Anton-Lamprecht I., Gjesti A.,  
RA Joergensen R.S., Jahnsen F.L., Olaisen B., Wiche G.,  
RA Gedde-Dahl T. Jr.;  
RT "A site-specific plectin mutation causes dominant epidermolysis  
RT bullosa simplex Ogna: two identical de novo mutations.";  
RL J. Invest. Dermatol. 118:87-93(2002).  
CC -!- FUNCTION: Interlinks intermediate filaments with microtubules and  
CC microfilaments and anchors intermediate filaments to desmosomes or  
CC hemidesmosomes. Could also bind muscle proteins such as actin to  
CC membrane complexes in muscle. May be involved not only in the  
CC crosslinking and stabilization of cytoskeletal intermediate  
CC filaments network, but also in the regulation of their dynamics.  
CC -!- SUBUNIT: Homodimer or homotetramer.  
CC -!- FUNCTION: Alternative splicing; Named isoforms=3;  
CC -!- ALTERNATIVE PRODUCTS:  
CC Name=1;  
CC IsoId=Q15149-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q15149-2; Sequence=VSP\_005030;  
CC Name=3;  
CC IsoId=Q15149-3; Sequence=VSP\_005030, VSP\_005031;  
CC TISSUE SPECIFICITY: Widely expressed with highest levels in  
CC muscle, heart, placenta and spinal cord.  
CC -!- DOMAIN: The N-terminus interacts with actin, the C-terminus with  
CC vimentin, desmin, GFAP, cytokeratins, lamin B; whereas both the N-  
CC and the C-terminus can bind integrin beta-4.  
CC -!- PTM: PHOSPHORYLATED BY CMC2; REGULATES DISSOCIATION FROM  
CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).  
CC -!- DISEASE: Defects in PLECI are the cause of epidermolysis bullosa  
CC simplex with muscular dystrophy (MD-EBS) [MIM:226670]; an  
CC autosomal recessive disorder characterized by epidermal blister  
CC formation at the level of the hemidesmosome and associated with  
CC late-onset muscular dystrophy.  
CC -!- DISEASE: Defects in PLECI are the cause of epidermolysis bullosa  
CC simplex 1 (EBS1) [MIM:131950]; also called epidermolysis bullosa  
CC simplex Ogna type. EBS1 is an autosomal dominant form of  
CC epidermolysis bullosa simplex differentiated from the more  
CC generalized form of Koebner [MIM:131900] and the localized form of  
CC Weber and Cockayne [MIM:131800] by the occurrence of skin  
CC bruising.  
CC -!- SIMILARITY: Contains 1 actin-binding domain.  
CC -!- SIMILARITY: Contains 2 calponin-homology (CH) domains.  
CC -!- SIMILARITY: Contains 33 plectrin repeats.  
CC -!- SIMILARITY: Contains 4 spectrin repeats.  
CC -!- SIMILARITY: Belongs to the plakin or cytolinker family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Z54367; CAA91196.1; --  
DR EMBL; U53204; AAB05427.1; --  
DR EMBL; U63610; AAB05428.1; --  
DR EMBL; U63609; AAB05428.1; JOINED.  
DR EMBL; X97053; CAA65765.1; --  
DR PIR; C59404; A59404.  
DR HSSP; Q01082; 1BKR.  
DR Genew; HGNC:9069; PLECI.  
DR GK; Q15149; --  
DR MIM; 601282; --  
DR MIM; 226670; --  
DR MIM; 131950; --  
DR GO; GO:0008307; F:structural constituent of muscle; TAS.  
DR InterPro; IPR001589; Actbind actnin.  
DR InterPro; IPR001715; Calponin-like.  
DR InterPro; IPR001101; Plectrin repeat.  
DR InterPro; IPR005326; S10\_plectrin\_N.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00307; CH; 2.  
DR Pfam; PF00681; Plectin; 19.  
DR Pfam; PF03501; S10\_plectin; 1.  
DR ProDom; PD006662; S10\_plectin\_N; 1.  
DR SMART; SM00033; CH; 2.  
DR SMART; SM00250; PLEC; 34.  
DR PROSITE; PS00019; ACTININ\_1; FALSE NEG.  
DR PROSITE; PS00020; ACTININ\_2; FALSE\_NEG.  
DR PROSITE; PS00021; CH; 2.  
DR Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;  
KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;  
KW Disease mutation.  
FT DOMAIN 1 1470 GLOBULAR 1.  
FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.  
FT DOMAIN 2756 4684 GLOBULAR 2.  
FT DOMAIN 175 400 ACTIN-BINDING.  
FT DOMAIN 179 282 CH 1.  
FT DOMAIN 295 397 CH 2.  
FT REPEAT 645 710 SPECTRIN 1.  
FT REPEAT 740 824 SPECTRIN 2.  
FT REPEAT 837 930 SPECTRIN 3.  
FT REPEAT 1315 1415 SPECTRIN 4.  
FT DOMAIN 1469 2756 COILED COIL (POTENTIAL).  
FT REPEAT 2826 2863 PLECTIN 1.  
FT REPEAT 2864 2901 PLECTIN 2.  
FT REPEAT 2902 2939 PLECTIN 3.  
FT REPEAT 2940 2977 PLECTIN 4.  
FT REPEAT 2981 3015 PLECTIN 5.  
FT REPEAT 3116 3153 PLECTIN 6.  
FT REPEAT 3154 3191 PLECTIN 7.  
FT REPEAT 3192 3229 PLECTIN 8.  
FT REPEAT 3230 3267 PLECTIN 9.  
FT REPEAT 3268 3305 PLECTIN 10.  
FT REPEAT 3306 3343 PLECTIN 11.  
FT REPEAT 3485 3522 PLECTIN 12.  
FT REPEAT 3523 3560 PLECTIN 13.  
FT REPEAT 3561 3598 PLECTIN 14.  
FT REPEAT 3599 3636 PLECTIN 15.  
FT REPEAT 3640 3674 PLECTIN 16.  
FT REPEAT 3820 3857 PLECTIN 17.  
FT REPEAT 3858 3895 PLECTIN 18.  
FT REPEAT 3896 3933 PLECTIN 19.  
FT REPEAT 3934 3971 PLECTIN 20.  
FT REPEAT 3975 4008 PLECTIN 21.  
FT REPEAT 4063 4100 PLECTIN 22.  
FT REPEAT 4101 4138 PLECTIN 23.  
FT REPEAT 4139 4176 PLECTIN 24.  
FT REPEAT 4177 4214 PLECTIN 25.  
FT REPEAT 4218 4252 PLECTIN 26.  
FT REPEAT 4265 4305 PLECTIN 27.  
FT REPEAT 4319 4356 PLECTIN 28.  
FT REPEAT 4408 4445 PLECTIN 29.  
FT REPEAT 4446 4483 PLECTIN 30.  
FT REPEAT 4484 4521 PLECTIN 31.  
FT REPEAT 4522 4559 PLECTIN 32.



CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

DR -----  
DR EMEL; AE001048; AAB90425.1; --  
DR PIR; A69352; A69352.  
DR TIGR; AF0817; --  
KW Hypothetical protein; Signal; Complete proteome.  
FT SIGNAL 1 21 POTENTIAL  
FT CHAIN 22 619 HYPOTHETICAL PROTEIN AF0817.  
SQ SEQUENCE 619 AA; 70225 MW; 8F13976711E7B24E CRC64;

Query Match 5.6%; Score 94.5; DB 1; Length 619;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 82; Conservative 50; Mismatches 109; Indels 141; Gaps 19;

QY	13	SIDTARALLKFFGDAQRYNVRSAVTLLALAGLK-----PGRWVD-----	53
Db	249	TVDTYTVWK---LDAKTGEVENSFVSGSMVVTMSKNAIYVTYNSYADPAKLTQF	305
QY	54	-STTP-----RLGVOKIMDWSGEHWAKPY-----ATGSR-----	81
Db	306	ISENPDLVPDWIREKIEKLMYDISSRAKQVEIMYLLEQLRASWSEDERLKFNENYNRW	365
QY	82	EDFRKKTILROWDNGFA---VLNADNLNIATNSQLNEYCLSDRALQALRAYGTEGFEES	137
Db	366	ENFTKKHARETEKTHIAKFSIQLEAEGMNSVPGRLNRFSL-DEYNGYLRVAITVDWDEN	424
QY	138	LVVFLDEASKAV-KARAEALQAMISVDLPGGEEFLLSPAG-----ONPLLKQKVE	187
Db	425	DLVVLDEKLEWVGKIQGFGLDERIYAVRFDGVDGFIYTRQTPFFVLDSLNPENPKIVG	484
QY	188	EF-VPRFA-----PRSTVLVYLGDTGR--KXSLFE-----REIFEVL--	221
Db	485	ELKIPGSSYLHRIDENTVLGVGREGNVKLSLFDISLTSPEKKNRYILQETWSEVLSN	544
QY	222	-----GLTFDPHGRMPDLILHDEVRGWLF-----LMEAVKSK-----	253
Db	545	HHAFLLDSQHGIFFLPAGQ-----NGYIFSYKDKGLKLIKAVKGNVRAIYIDDL	594
QY	254	---GP-----FDEERHRSLOEL	267
Db	595	YIIGPEEISYDENSWEKVGEL	616

Search completed: October 1, 2004, 16:13:21  
Job time : 10.5859 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 31.5414 Seconds  
(without alignments)  
3231.066 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSGDIGTVASIDTARAL.....EDPHLIHLNGSRFLGPFYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_25.\*

1: sp.archaea.\*

2: sp.bacteria.\*

3: sp.fungi.\*

4: sp.human.\*

5: sp.invertebrate.\*

6: sp.mammal.\*

7: sp.mhc.\*

8: sp.organelle.\*

9: sp.phage.\*

10: sp.plant.\*

11: sp.rodent.\*

12: sp.virus.\*

13: sp.vertibrate.\*

14: sp.unclassified.\*

15: sp.rvirus.\*

16: sp.bacteriap.\*

17: sp.archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match	Length	DB	ID	Description
1	632.5	37.4	546	16	Q8YQW1		Q8YQW1 anabaena sp
2	604.5	35.8	826	2	Q8VLS0		Q8VLS0 yersinia en
3	240.5	14.2	364	2	Q9RB06		Q9RB06 xanthomonas
4	207	12.3	353	2	Q52863		Q52863 xanthomonas
5	118	7.0	1111	16	Q8NLY2		Q8NLY2 corynebacte
6	103	6.1	331	11	Q9JL56		Q9JL56 ratius norv
7	102.5	6.1	444	16	Q819U2		Q819U2 bacillus ce
8	102.5	6.1	671	16	Q8EQH8		Q8EQH8 oceanobacil
9	102	6.0	496	2	Q7X5B0		Q7X5B0 fischerella
10	102	6.0	501	16	Q7V954		Q7V954 prochloroco
11	101.5	6.0	556	12	Q7TFK6		Q7TFK6 rhesus cyto
12	101.5	6.0	1057	16	Q8CPJ4		Q8CPJ4 staphylococ
13	100.5	6.0	459	16	Q8E409		Q8E409 streptococc
14	100	5.9	498	15	Q77372		Q77372 human immun
15	100	5.9	772	12	Q06907		Q06907 bovine rota
16	99.5	5.9	435	2	Q9F0V8		Q9F0V8 azoarcus sp

17	99.5	5.9	1203	16	Q9A2L9		Q9A2L9 caulobacter
18	99	5.9	496	2	Q47918		Q47918 fischerella
19	99	5.9	496	2	Q8GPM2		Q8GPM2 scytonema h
20	99	5.9	496	2	Q8GPM1		Q8GPM1 fischerella
21	99	5.9	496	2	Q7X5B1		Q7X5B1 scytonema s
22	99	5.9	496	2	Q7X5A7		Q7X5A7 fischerella
23	98.5	5.8	331	11	Q9JL56		Q9JL56 mus musculu
24	98.5	5.8	838	3	Q871I3		Q871I3 neurospora
25	97.5	5.8	1346	16	Q98IR5		Q98IR5 rhizobium l
26	97	5.7	772	12	Q8QPV9		Q8QPV9 bovine rota
27	96	5.7	364	2	Q8GDI7		Q8GDI7 heliobacill
28	96	5.7	496	2	Q8GPN0		Q8GPN0 calothrix d
29	96	5.7	527	16	Q8XZ67		Q8XZ67 ralstonia s
30	95	5.7	793	11	Q8C2S1		Q8C2S1 mus musculu
31	95.5	5.7	428	16	Q8KES9		Q8KES9 chlorobium
32	95.5	5.7	456	2	Q9EY17		Q9EY17 salmonella
33	95.5	5.7	497	15	Q900Z2		Q900Z2 human immun
34	95.5	5.7	797	16	Q915T8		Q915T8 pseudomonas
35	95	5.6	389	17	Q971H8		Q971H8 sulfobus
36	95	5.6	409	16	Q8U786		Q8U786 agrobacteri
37	95	5.6	497	15	Q901A4		Q901A4 human immun
38	94.5	5.6	391	10	Q8GS76		Q8GS76 oryza sativ
39	94.5	5.6	456	2	Q9EY8		Q9EY8 salmonella
40	94.5	5.6	574	5	O01939		O01939 drosophila
41	94.5	5.6	1445	15	Q900Z1		Q900Z1 human immun
42	94	5.6	155	16	Q8NNE1		Q8NNE1 corynebacte
43	94	5.6	284	16	Q8YH18		Q8YH18 brucella me
44	94	5.6	288	16	Q8G0C2		Q8G0C2 brucella su
45	94	5.6	312	16	P74729		P74729 synecocyst

#### ALIGNMENTS

#### RESULT 1

Q8YQW1 ID Q8YQW1 PRELIMINARY; PRT; 546 AA.  
AC Q8YQW1;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Type II site-specific deoxyribonuclease.  
GN ALR3701.  
OS Anabaena sp. (strain PCC 7120).  
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
OX NCBI\_TaxID=103690;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21595285; PubMed=11759840;  
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
RA Yasuda M., Tabata S.;  
RT "Complete genomic sequence of the filamentous nitrogen-fixing  
RT cyanobacterium Anabaena sp. strain PCC 7120.";  
RL DNA Res. 8:205-213(2001).  
DR EMBL; AF003594; BAB75400.1; -.  
DR PIR; AF2268; AF2268.  
DR Complete proteome.  
SQ SEQUENCE 546 AA; 62662 MW; BCA9A4C3CCB2A5E8 CRC64;

Query Match 37.4%; Score 632.5; DB 16; Length 546;  
Best Local Similarity 40.2%; Pred. No. 1.9e-43;  
Matches 129; Conservative 65; Mismatches 122; Indels 5; Gaps 3;  
QY 4 SDGIDGTVAS--IDTARALLKRFDFDAQYNYRSVATLLALACLKPGDRVNSTTRELGV 61  
Db 223 SBAIKAVLAARKEIEALAILKDISAPKQONERSALFLALADIRPTEIWTQATSPRRRI 282  
QY 62 QKIMDSGSHWAKPYATGSRDFRKKTKLQWVDNGFAVLNADNLNATNSQLNEYCISDE 121  
Db 283 TEMDMWFRDHYGKQVAPNTRVTRQTMQFQVMGIVVENPQDPDPINSPKWCYQLHQQ 342

```
QY 122 ALQALRAYGTGFPRESLVVFLDEASKAVKARAEALQAAMISVDLPGGEEFLLSPAGQNPL 181
Db 343 FVTLLKSSQWEETRNNYVISVKNLLQDRNNI--PMIPVSPNGQAIQLSSGGQNIL 400
QY 182 LKKWVEEFVPRFAPRSTVLYLGDTRGKHSLEPERIFEIVLGLTDPHGMPLDILHDEVR 241
Db 401 IKEILENFCPRFTPEGVLVFGDAGNKFIVNETQKFEI-GIELDPHGKMPDIVVYERQ 459
QY 242 GWLFIMEAVKSGPPDEERHRSLOELFVTPSAGLIFVNCFNRRSMRWLPELAWETAW 301
Db 460 EWLVLIEAVTGHGPNLKRRELKLFQSSRQGLVFTVTAFFSRKEMTRYLAELISMETEVM 519
QY 302 VAEDPDHLHLNGSRFLGPE 322
Db 520 VAAQPDHMHFNGERFLGPE 540

RESULT 2
Q8VL50 PRELIMINARY; PRT; 826 AA.
ID Q8VL50
AC Q8VL50;
DT 01-MAR-2002 (TremBLrel. 20, Created)
DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE Methyltransferase-endonuclease.
GN YENL.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM=IS1222;
RA Rakin A.V.;
RT "Characterization of the restriction-modification system of Y.
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
EMBL; AJ414030; CAC95150.1; --
DB GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004519; F:endonuclease activity; IEA.
DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
DR GO; GO:0008757; F:S-adenosylmethionine-dependent methyltransf. . ; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR InterPro; IPR002296; N12N6_mtfase.
DR InterPro; IPR002052; N6_Mtase.
DR PRINTS; PR000051; SAM bind.
DR PROSITE; PS00092; N6_MTASE; 1.
KW Transferase; Endonuclease.
SQ SEQUENCE 826 AA; 93348 MW; 148432FD8E798843 CRC64;

Query Match 35.8%; Score 604.5; DB 2; Length 826;
Best Local Similarity 39.3%; Pred. No. 7.1e-41;
Matches 121; Conservative 58; Mismatches 128; Indels 1; Gaps 1;

QY 14 IDTARALKRFGDAQRNVRSAVILLALAGLKPGDRWVDSSTPRLGVQKIMDNGEHW 73
Db 519 LDQALEIITLLGMPSQQNERSALTFALVNLRPESGWOELKPLGVGTPIMDCRDIYG 578
QY 74 KYATGSRDFRKKTLROWDNGFVNLNADNLIATNSQLNECYCLSDALQALRAYGTG 133
Db 579 KEYANTRETFRQTLHQFIDGLVLYNPKENRNVNSPKACYQIAPELFDVLYNTYGTPL 638
QY 134 FRESLVVFLDEASKAVKARAEALQAAMISVDLPGGEEFLLSPAGQNPLKKWVEFVPRF 193
Db 639 WNKALGEWLWQRETLVEQYAMKREMHMIIPLTIDNGTEIHLSPGDHSLIHDIVTEFGPR 698
QY 194 APRSTVLYLGDTRGKHSLEPERIFEIVLGLTDPHGMPLDILHDEVRGWLFIWEAVKSK 253
Db 699 APGSQVILYLGDTAKEDDFRKKDALAD-LGVTVNRKGLPDVVLWYFQDNLILIESVTSH 757
```

```
QY 254 GFPEERHRSLOELFVTPSAGLIFVNCFNRRSMRWLPELAWETAWVADDPDHLIHLN 313
Db 758 GPVDGKRHSLEALFKDARPGLVVVSAPPDKTKWSPFSEISWETVWIAEAPTHMIHLN 817
QY 314 GSRLFGPY 321
Db 818 GDRELGPH 825

RESULT 3
Q9RBJ6 PRELIMINARY; PRT; 364 AA.
ID Q9RBJ6
AC Q9RBJ6;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
DE Restriction endonuclease homolog R.XphI.
GN XPHIR.
OS Xanthomonas campestris (pv. phaseoli).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=29445;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Xcp 73;
RA Lai J.-Y., Yang M.-T.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF042157; AAF22367.1; --
DR GO; GO:0004519; F:endonuclease activity; IEA.
KW Endonuclease.
SQ SEQUENCE 364 AA; 39725 MW; 8D150D6F40069A3D CRC64;

Query Match 14.2%; Score 240.5; DB 2; Length 364;
Best Local Similarity 27.2%; Pred. No. 2.1e-11;
Matches 94; Conservative 56; Mismatches 118; Indels 77; Gaps 18;

QY 7 IDGT-----VASIDTARALLKRFQDAQRNVRSAVTLALAGLKPGDRWVDSSTPR 58
Db 48 VEGTGRVLAPKHVMRTVEALLSD---DA---VREA---YGLGAMKPGFR----- 89
QY 59 LGVQKIMDNGEHWAKPYATGSRDFRKKTLROWDNGFAVLN--ADNLIATNSQLNEY 116
Db 90 -----VAGQRW---YEENSREPLRDLRQ-----GFITNNAVAERTGLFTTGLPRY 134
QY 117 CL-----SDEALQALRAYGTGFEESLVVFLDEASKAVKARAEALQAAMISV 163
Db 135 ALKTDFAALFDPALAGDILLAIITAWQEAHLASAM-----ARIALVRGAAPTDEGVW 189
QY 164 DLPGGEEFLLSPAGQNPLKKWVEFVPRFAPRSTVLYLGDTRGKHSLEPERIFEIV-IG 222
Db 190 TFPNGETRRMAPGPSVISKAVIEEFAARFLTQPAVLWVSESAKVVSRRDDELAKSLK 249
QY 223 LTFDPHGMPLDILHD-----EVRGWLFL-MEAVKSGPDEERHRSLOELFV----TSA 273
Db 250 ITADRN--LPDILVLDGGGVTFGLLVFIEVATDGPITMQRREAFWQIAADAGFTTPEA 307
QY 274 GLIFVNCFNRR--SMRWLPELAWETAWVADDPDHLIHL-NGS 315
Db 308 -VAFVATYLDRSHTAFKKTIAELANRSEFAWFASEPEHIIALHNGA 351

RESULT 4
Q52863 PRELIMINARY; PRT; 353 AA.
ID Q52863
AC Q52863;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Endonuclease.
OS Rhizobium leguminosarum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
OX NCBI_TaxID=384;
```

```

[1]
SEQUENCE FROM N.A.
RX MEDLINE=98055154; PubMed=9393436;
RX Rochepeau P., Selinger L.B., Hynes M.F.;
RT "Transposon-like structure of a new plasmid encoded restriction-
RT modification system in Rhizobium leguminosarum VF39SM.";
RT Mol. Gen. Genet. 256:387-396(1997).
DR EMBL; X99520; CAA67875.1; -.
SQ SEQUENCE 353 AA; 38174 MW; C74CF7DC8D011A7B CRC64;

Query Match          12.3%; Score 207; DB 2; Length 353;
Best Local Similarity 24.0%; Pred. No. 1.2e-08;
Matches 82; Conservative 51; Mismatches 145; Indels 64; Gaps 12;

QY 1 VNSDGDIGGVASIDTARALLKRGFGDAQRVNVRSVTLALAGLKGCDRMVDSTTPRLG 60
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 46 INAVDGGIQLAPKHVYVMTQEQSLTDAHRLAYATAIMKAGHVDERW----- 96
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 61 VQKIMDWSGEHWAKPYATGSRDPKKTLROWVDNGFAVLNADNLNATNSQLNEY---- 116
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 97 -----YLDNTRRESIRDDTSREALVATGAVI--EDTTAATTSKGKRYALQA 139
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 117 -----CLSDREAL-QAALR-----AYTEGFEESLVFLDEASKAVKARABALQAAMISV 163
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 140 AFAALMVPALQDEALDQAINENWGRYITAGALARVAI-----VRAGA-ATGCTHVQV 190
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 164 DLPGBEEFLSPAGONPLLKQWVEFVPAPRSTVLVGLDTRKGKHSIFREEIFEVLGL 223
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 191 TFFNGETRLKAGESPSSVITKDVIEVFSRFLGDPFAVLVFSSES-ONKVVARDEKLTATIGL 249
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 224 TFDPHGRMPDLILHD--EVRGWLFLMEAVKSGPFDDEHRHSLOELFVFTPSAG-----LI 276
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 250 SIQSDKDLDPDIIILVDLPAKPELLVFVEVATDGPVGVRRAALEK--LAEDAGFDLQHVA 307
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

QY 277 FVNCFNENR--ESMRQWLPELAWETEAVVADPDHLIHLNGSR 316
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 308 FVTAYLDRSOSTFKRTAETLAWGTFAWFAGPEHIVELSEGR 349
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::

```

```

RESULT 5
Q8NLY2
ID Q8NLY2 PRELIMINARY; PRT; 1111 AA.
AC Q8NLY2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein Cgl2804.
DE CGL2804.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005283; BAC00198.1; -.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 1111 AA; 123962 MW; E354FB360EDDF9A CRC64;

```

```
Query Match      7.0%; Score 118; DB 16; Length 1111;  
Best Local Similarity 25.5%; Pred. No. 1.4;  
Matches        60; Conservative 27; Mismatches 104; Indels 44; Gaps    7;
```

-----

Qy	29	QRNVRSAVTLALAGLIKPGDRWVDSSTPRIGVKIMDWSGE---HWAK-----P	75
	:	: ::  :	:
Dd	570	KRFNIRCVTRPEELSSALGPRDQGVTI----LGVKFAQQTGDPTRWEKDRRKLGRDST	625
	:	: ::  :	:
Qy	76	YATGSREDFRKKTIROWVDNGFAYLN-ADNLNIATNSQLNE-----YCLSEALQALRAYG	130
	:	: ::  :	:
Dd	626	YRLGSTNDAAKVETLRFTFKAGVAVVQAADRNRIAANRAELRELERQYOASOEILIKVSWAQI	685
	:	: ::  :	:

Qy	131	TGTFEESUUVLD-----EASKAVKARAEALQAAMISVDLPGEFF	171
Dd	686	DVESADAIAELDRLELNNTPEATLSARHEAAKQTLRVSDLIIVAAQSSETVASML	745
Qy	172	LLSPAGQNPLIKKKVEEPVPFAPRSTVLYLGDTRGHSLPERFEFVLGTFD	226
Dd	746	KRAETKLUESLPAEVSEIAREVEKLFPLANTRVHA--ANVDEQTIALRED	797
RESULT 6			
ID	Q9JUL5	PRELIMINARY; PRT: 331 AA.	
AC	Q9JUL5;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last annotation update)		
DE	Membrane interacting protein of RGS16.		
GN	MIR16.		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxID=10116;		
EN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20226047; PubMed=10760272;		
RA	Zhang B., Chen D., Farquhar M.G.;		
RT	"MIR16, a putative membrane glycerophosphodiester phosphodiesterase,		
RT	interacts with RGS16.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 97:3999-4004(2000) .		
DR	EMBL: AF212861; AAF65233.1;		
DR	GO; GO:0008899; F:glycerophosphodiester phosphodiesterase act. . ; IEA.		
DR	GO; GO:0006071; P:glycerol metabolism; IEA.		
DR	InterPro; IPR004129; GDDP.		
SQ	Pfam; PF03009; GDDP; 1.		
SQ	SEQUENCE 331 AA; 37633 MW; D946067B5766E35F CRC64;		

Query Match	6.1%;	Score 103;	DB 11;	Length 331;	
Best Local Similarity	23.0%;	Pred. No. 3.9;			
Matches	78;	Conservative 36;	Mismatches 95;	Indels 130;	Gaps 19
Qy	20	LLKRFGDAQRYNVSRAVTLLALAGLKPGDRWDVSTTPRLGVQKIMDWSGEH-----	71		
Db	41	LLRFFSFE-----PVPSPRALQVLKPRDR-----VSAIAHRGGSHDAPENTLTA	83		
Qy	72	-----WAKPYATGSRREDPRFKKTLRQWDNGFAVLNADN-----	105		
Db	84	AIROAAKNGATGVELDI-----EFTSDGVPVLMHDNTVDRTTDGSGRGLCDLTFBQVRKL	137		
Qy	106	NIATNSQL-NEYCLSDREALQALRAYGTGFPESLVVFLDDEASKAVKAAEALQAAMISVD	164		
Db	138	NPAANHRLREF--PDRIPRTLREAVTECLCHNLTIFD-----VKGHADMASAAAKNIY	190		
Qy	165	LPGGEFEELLSPAGQNPLL--KKMVEEFVPR--FAPRST-----VLVLGDTR	206		
Db	191	M-----EF-----POLYNNSMVCSFLPEVIYKMRQTDQKVITALTHRPWSLSHTGDGK	238		
Qy	207	GKHSLFEREIPEEVGLGTFD--PHG-----RMPDLILHDEVGRVIMFLMEAVKS	252		
Db	239	PRYSVFWKQSVFVVLIDLLDSMENVLMWYLCISAFILMKQDFVSPDYLLKKW-----SA	291		
Qy	253	KG-----PPDEERH--RSLQELFVTPSAGLIFVNC	280		
Db	292	KGIQVSWTNTFPDEKNYYESHLGSSYITDS---MLENC	327		
RESULT 7					
Q819U2					
ID	Q819U2	PRELIMINARY;	PRT;	444	AA.
AC	Q819U2;				
DT	01-JUN-2003	(TrEMBLrel. 24, Created)			
DT	01-JUN-2003	(TrEMBLrel. 24, Last sequence update)			
DT	01-OCT-2003	(TrEMBLrel. 25, Last annotation update)			





```
GN SE0879,
OS Staphylococcus epidermidis
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren Y., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016746; AA004476.1; -
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004086; F:carbamoyl-phosphate synthase activity; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0006526; P:arginine biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0019856; P:pyrimidine base biosynthesis; IEA.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; Cphase_L.
DR InterPro; IPR005479; Cphase_L_D2.
DR InterPro; IPR005480; Cphase_L_D3.
DR InterPro; IPR005481; Cphase_L_N.
DR InterPro; IPR004362; MGS-like.
DR InterPro; IPR000169; SproT_acsite.
DR Pfam; PF00289; CPSase_L.
DR Pfam; PF02786; CPSase_L_D2; 2.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR00098; CPSASE.
DR TIGRFAMs; TIGR01369; CPSaseII_lrg; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; 1.
DR Complete proteome.
SQ SEQUENCE 1057 AA; 117391 MW; 8944D7D8DBICAE59 CRC64;

Query Match 6.0%; Score 101.5; DB 16; Length 1057;
Best Local Similarity 22.7%; Pred. No. 30;
Matches 61; Conservative 40; Mismatches 83; Indels 85; Gaps 14;

QY 82 EDFRKTLQWVDNGFAVLNADNL-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIYEVRNDRKNDNAIVVCNENIDPVGIHGTDSIVVAPSTLSVETQMRLDVS�KVR 273
QY 128 AYGTEG-----FEESLVVFLDEASKAVKARAEALQ-----AAMISVDLPFG 168
Db 274 ALGIEGGCNVQLADPHSLNYVILVNPV-SSSALASKATGYPIAKLAIAVGLTLD 332
QY 169 EEFLLSP-----AQCNPLKKMVEFVPR-----APRSTVLYLGDT 206
Db 333 E--MLNPITGTSYAAFEPTLDYVISK-IPRFPDFKFKGERELGTQMKATGEWMAIGRT- 388
QY 207 GKHSLFEREIPEVLGLTFDHP-----GRWPDL-----ILHDSVRGWLFLMEAVKSKG 254
Db 389 -----YEESLKAINSLRYGVHHLGSLNGESVELDYIKERIGHQDDRLFFIGEAIR--- 440
QY 255 PFDEERHRSQELFVTPSAGLIFVNCFN 283
Db 441 -----RGTSLEELHNMTKIDYFFLNKFQN 464

RESULT 13
Q8E409
ID Q8E409 PRELIMINARY; PRT; 459 AA.
AC Q8E409;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN GBS1594.

OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX NCBI_TaxID=216495;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEM316 / Serotype III;
RA MEDLINE=22242508; PubMed=12354221;
RA Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Msadek T., Zouine M., Couve E., Lalioui L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RA "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766852; CA047253.1; -
DR Sagallst; gbs1594; -
DR GO; GO:0003977; F:UDP-N-acetylglucosamine diphosphorylase act. .; IEA.
DR GO; GO:0009058; P:biosynthesis; IEA.
DR GO; GO:0009103; P:lipopolysaccharide biosynthesis; IEA.
DR InterPro; IPR005882; GlmU.
DR InterPro; IPR001451; Hexapep transf.
DR InterPro; IPR005835; NTP transferase.
DR Pfam; PF00132; hexapep; 7.
DR Pfam; PF00483; NTP transferase; 1.
DR TIGRFAMs; TIGR01173; GlmU; 1.
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 459 AA; 49377 MW; 4D6FE3E9E4629658 CRC64;

Query Match 6.0%; Score 100.5; DB 16; Length 459;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 60; Conservative 37; Mismatches 94; Indels 59; Gaps 13;

QY 16 TARALLKRFQDAQRYNVRSVATLLALAGLXPGDRWVDSTTPRLGVOKIMDWSGSHWAKP 75
Db 107 TGESLNLIHPHVNKKVATILT-----ADANP-FGYGRIIRNSDDEVTKI 152
QY 76 YATGSRDFRKTKLRQWVDNGFAVLNADNL-----NIATNSQLNVEYCLSD-----EALQ 124
Db 153 VEQKDANDFEQV--KEINTGTYYFDNQSLFEALKDINTNNAQGEYLTLDVIGIFKEAGK 210
QY 125 ALRAYGTGFPESL-----VFLDEASKAVK---ARAEL-----QAAMISVDLPGE 169
Db 211 KVGAYKLRDFDESIGVNDRVALATAEKVMHRIARQHMVNGVTVVNPDSAYIDIDVEIGE 270
QY 170 EFLLSAQCNPLKKMVEFVPRFAPRSTVL-----YLGDTL-RGKHSLFEREIPEVL--- 221
Db 271 ESVEIP---NVLKGQTK-----IGKTLTNGSYLVDQVGNVDVTITNSMVEESIISD 321
QY 222 GLTFDPHGMR 231
Db 322 GVTGVPYAH 331

RESULT 14
Q77372
ID Q77372 PRELIMINARY; PRT; 498 AA.
AC Q77372;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Gag polyprotein.
GN GAG.
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=94149849; PubMed=8107220;
RA Vanden Haesevelde M., Decourt J.L., De Leys R.J., Vanderborght B.,
RA van der Groen G., van Heuverswijn H., Saman E.;
RT "Genomic cloning and complete sequence analysis of a highly divergent
RT African human immunodeficiency virus isolate.";
RN J. Virol. 68:1586-1596(1994).
```

```
DR EMBL; L20587; AAA99878.1; -.
DR HSP; P05888; IAAF.
DR GO; GO:0019012; C:virion; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR GO; GO:0016032; P:viral life cycle; IEA.
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR000071; Retroviral p17.
DR InterPro; IPR008916; Retrov capsid C.
DR InterPro; IPR008919; Retrov capsid N.
DR InterPro; IPR001878; Znf CCHC.
DR Pfam; PF00540; Gag_p17; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; zf-CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR PRINTS; PR00234; HIV1MATRIX.
DR SMART; SM00343; Znf C2HC; 2.
DR PROSITE; PS00158; ZF_CCHC; 2.
KW AIDS; Core protein; Polyprotein.
SQ SEQUENCE 498 AA; 55304 MW; 64D16C9542519693 CRC64;

Query Match          5.9%; Score 100; DB 15; Length 498;
Best Local Similarity 24.6%; Pred. No. 13;
Matches 65; Conservative 33; Mismatches 100; Indels 66; Gaps 14;

QY 48 GD---RWVDSSTPRLGVQKIMDSGEHWAKPYATGSGREDPR---KKTLRQWDNGFAVLN 101
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 258 GDYRKWI-----VLGLNKMV-----KWSFVSILDIKQGPFRDYVDRFYKTLR 304

QY 102 ANLNLTATNSQLNEYCL-----SDEALQALRAYTEG-FEESLVVFLDEASKAVKARAEAL 156
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 305 AEQATQEVKNMTETILLVQNPANPDCKQLKSLGPGATLEEMVACQGVGPTHKARVLAE 364

QY 157 QAAMISVDLPGGSEFLLSPAGQNPLKKMVEEF-----VPR--FAPRSTVLY----- 201
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 365 ANATAQDLKGGTVAVFMORGQNPFRKGTIKCNCKGEGHIANRCRAPRKKGCWKQGEQ 424

QY 202 --LGDTG-KGKSLFEREIEFEVLGLTFDPHGRMPDLI-----LHDEVRGWLFML 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 HQMKDCRNGKQANF-----LGKWPFGGTREPGNVQRPAPSPAPMEEVKG---Q 472

QY 248 EAVKSK-GPFDEERHRSLOELFVT 270
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 473 ENQEQKGGPNELYPFASLKSFGT 496

RESULT 15
ID Q06907 PRELIMINARY; PRT; 772 AA.
AC Q06907;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE VP4 protein.
GN VP4.
OS Bovine rotavirus.
OC Viruses; dsRNA viruses; Reoviridae; Rotavirus.
OX NCBI_TaxID=10927;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=KK-3;
RX MEDLINE=94049850; PubMed=8232344;
RA Isegawa Y., Nakagomi O., Nakagomi T., Ishida S., Uesugi S., Ueda S.;
RT "Determination of bovine rotavirus G and P serotypes by polymerase
chain reaction.";
RL Mol. Cell. Probes 7:277-284 (1993).
DR EMBL; D14367; BAA03284.1; -.
DR GO; GO:0019028; C:viral capsid; IEA.
DR InterPro; IPR000416; Cap VP4.
DR InterPro; IPR008985; ConA_like_lec_gl.
DR Pfam; PF00426; VP4; 1.
SQ SEQUENCE 772 AA; 86724 MW; E8A29D8B89EF9F15 CRC64;
```

```
Query Match          5.9%; Score 100; DB 12; Length 772;
Best Local Similarity 22.9%; Pred. No. 25;
Matches 77; Conservative 45; Mismatches 130; Indels 84; Gaps 16;

QY 1 VNSDGDIGTVASIDTARALLKRFDFDAQRYNVRSVAVTLALA-----GLKPGDRWYDST 55
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 473 VPSNDDYQTPIANSVTVRQDLER-QLDEMKEFNLSANIALSLQIDLALLP---LDMF 527

QY 56 TPRLGQVKIMDSGEHWAKPYATGSGREDPRK-----KTLRQ 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 528 SMFSGIQSTVE-----AAKSFATSVMKFKRSLAKSVNSLTDAITDAAGSISRSSTLRS 582

QY 92 -----WVDNGFAVLNADNLNLTATNSQLNEYCLSDLEALQALRAYTEGF--RESLVFL 142
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 583 VNSAVSVWTDISDIVDSTDNVVAAT-----ATAAKKFRVKEFTTFDGVGSFD 630

QY 143 DEASKAVKARAEALQAAMISVDLPGGSEFLLSPAGQNPLKKMVEEFVPRFAPR---STV 199
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 631 DISAAVVVKTKMKNLNV---VD---EEL-----PQIITEASEKFIPIPNRAYRLIDGEX 676

QY 200 LYLGDTRGKHSLEFERIEFEVLGLTFDPHGRMPDLILHDEVRGWLFMEAVKS-KGPFDE 258
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 677 VVEVTTEGKYFAYLTTFEEVW---FDAE-RFAELVTDSPVISAIIIDFKTIKNLDNYGI 732

QY 259 ERHRSLOELFVTSPAGLIIFVNCFENRESMRQWLPEL 294
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 733 TREQALNMLRSDPKVLRSFIN--QNNPIIKNRIEQL 766
```

Search completed: October 1, 2004, 16:10:31  
Job time : 35.5414 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:08:57 ; Search time 42.3214 Seconds  
(without alignments)  
2156.424 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIQGTWASIDTARAL.....EDPDHLHLNGSRFLGPYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A\_Geneseq\_29Jan04.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	118	7.0	708	4 AAB76666	Aab76666 Coryneb
2	118	7.0	713	4 AAB76665	Aab76665 Coryneb
3	118	7.0	1111	4 AAG92827	Aag92827 C glutami
4	105	6.2	135	6 ABM69452	Abm69452 Phototrab
5	103.5	6.1	1037	5 ABP39949	Abp39949 Staphyloc
6	103.5	6.1	1120	6 ABU42979	Abu42979 Protein e
7	100	5.9	772	6 Aaw57632	Aaw57632 VP4 prote
8	99.5	5.9	1057	6 ABU43956	Abu43956 Protein e
9	97.5	5.8	289	4 AAU36347	Aau36347 Pseudomon
10	97.5	5.8	289	6 ABU38597	Abu38597 Protein e
11	96.5	5.7	513	2 AAY05897	Aay05897 Vicia sat
12	95.5	5.7	797	4 AAU33575	Aau33575 Pseudomon
13	95.5	5.7	797	6 ABU15621	Abu15621 Protein e
14	95.5	5.7	1057	6 ABJ18981	Abj18981 Pathogen
15	95.5	5.7	1057	6 ABU42381	Abu42381 Protein e
16	95.5	5.7	1057	6 ABM71258	Abm71258 Staphyloc
17	94.5	5.6	574	4 ABB58101	Abb58101 Drosophil
18	94	5.6	155	4 AAG92225	Aag92225 C glutami
19	93.5	5.5	459	5 ABP26548	Abp26548 Streptoco
20	93	5.5	498	2 AAR51691	Aar51691 HIV-type
21	93	5.5	438	2 AAN33076	Aan33076 HIV isola
22	93	5.5	1080	4 AAM00803	Aam00803 Human bon
23	93	5.5	1637	4 AAM00916	Aam00916 Human bon
24	93	5.5	2263	4 AAM79000	Aam79000 Human pro
25	92.5	5.5	453	4 AAB96341	Aab96341 Putative

## ALIGNMENTS

### RESULT 1

AAB76666  
ID AAB76666 standard; protein; 708 AA.

XX AC

XX AAB76666;

XX AC

XX 11-APR-2001 (first entry)

XX XX

XX Corynebacterium glutamicum MCT protein SEQ ID NO:314.

XX XX

XX Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;

XX KW membrane construction and membrane transport protein; petroleum spill;

XX KW hydrocarbon degradation; gram positive aerobic bacterium; marker;

XX KW identification; microorganism; fine chemical production; transformation;

XX KW genome mapping; genetic engineering.

XX XX

XX Corynebacterium glutamicum.

XX OS

XX WO200100805-A2.

XX XX

XX 04-JAN-2001.

XX PD

XX 23-JUN-2000; 2000WO-IB000926.

XX XX

XX 25-JUN-1999; 99US-0141031P.

XX PR

XX 08-JUL-1999; 99DE-01031454.

XX PR

XX 08-JUL-1999; 99DE-01031478.

XX PR

XX 08-JUL-1999; 99DE-01031563.

XX PR

XX 09-JUL-1999; 99DE-01032122.

XX PR

XX 09-JUL-1999; 99DE-01032124.

XX PR

XX 03-JUL-1999; 99DE-01032125.

XX PR

XX 09-JUL-1999; 99DE-01032128.

XX PR

XX 09-JUL-1999; 99DE-01032180.

XX PR

XX 09-JUL-1999; 99DE-01032182.

XX PR

XX 09-JUL-1999; 99DE-01032190.

XX PR

XX 09-JUL-1999; 99DE-01032191.

XX PR

XX 09-JUL-1999; 99DE-01032209.

XX PR

XX 09-JUL-1999; 99DE-01032212.

XX PR

XX 09-JUL-1999; 99DE-01032227.

XX PR

XX 09-JUL-1999; 99DE-01032228.

Aau33514 Enterococ  
Aau33515 Enterococ  
Abu14587 Protein e  
Adu141645 Human DIT  
Ade29065 Human AIM  
Abg74680 Human CGD  
Abu17595 Protein e  
Abp70393 Amino aci  
Aab85856 Human MSH  
Aag63958 Amino aci  
Aao18557 Human mis  
Abu07465 Protein d  
Abu07976 Human MSH  
Abu89664 Human MSH  
Aao27522 Human mis  
Ada06254 Human mis  
Adc89615 Human MSH  
Aab12559 Coprinus  
Aab12558 Coprinus  
Abb93598 Herbicida

```
PR 27-AUG-1999; 99DE-01040830.
PR 27-AUG-1999; 99DE-01040831.
PR 27-AUG-1999; 99DE-01040832.
PR 27-AUG-1999; 99DE-01040833.
PR 31-AUG-1999; 99DE-01041378.
PR 31-AUG-1999; 99DE-01041379.
PR 31-AUG-1999; 99DE-01041395.
PR 03-SEP-1999; 99DE-01042077.
PR 03-SEP-1999; 99DE-01042078.
PR 03-SEP-1999; 99DE-01042079.
PR 03-SEP-1999; 99DE-01042088.
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-071486/08.
XX N-PSDB; AAF67899.
XX
Corynebacterium glutamicum nucleic acids encoding membrane construction
and membrane transport proteins or their portions, useful for typing or
identifying C. glutamicum or related bacteria, and as markers for
transformation.
XX
XX Claim 20; Page 617-619; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
construction and membrane transport (MCT) proteins given in AAB76510 to
AAB76847. The MCT nucleic acids and proteins are useful in the
identification of microorganisms which can be used to produce fine
chemicals, for modulating fine chemical production in C. glutamicum or
related bacteria (e.g. Brevibacterium lactofermentum), the typing or
identification of C. glutamicum or related bacteria, as reference points
for mapping C. glutamicum genome, and as markers for transformation.
XX
XX AAF68082 and AAF68082 represent sequencing primers which are used in an
example from the present invention
XX
XX Sequence 708 AA;
XX
Query Match 7.0%; Score 118; DB 4; Length 708;
Best Local Similarity 25.5%; Pred. No. 0.026;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;
Qy 29 QRYNRSVATLLALAGLPGDRWDSTTPRLGVQKIMDWSGE---HWAK-----P 75
Db 167 KRFNRCVTPBELSALGRDQGTI----LGVRFQAQGTGPTTRWEKDKRLGDRST 222
Qy 76 YATGSREDFRKTLRQWDNGFAVLN-ADNLNIATNSQLNE----YCLSDALQALRAYG 130
Db 223 YRLGSTNDKAVETLRETQKAGKAVQAADNRILAAANRAELRELERQYQASQELTKVSWAQI 282
Qy 131 TEGFPESLVVFLD-----EASKVKARAEALQAAMISVDLPGGREF 171
Db 293 DVESAATAIAELDRILLELNNTPEATLSARHEAAKQTLARVSDLLVAQSEETVASMNL 342
Qy 172 LLSPAGQNPLLKMWVEEFVPRFAPRSTVLVLDGTRGKSLPREFEVLGLTFD 226
Db 343 KRAETELKRLSLPVAEVSSEETAREVEKFLANTRVHA---ANVDEQTIALRED 394
RESULT 2
AAB76665
ID AAB76665 standard; protein; 713 AA.
XX
XX AAB76665;
XX
XX 11-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MCT protein SEQ ID NO:312.
XX
XX Corynebacterium glutamicum; Brevibacterium lactofermentum; MCT;
XX membrane construction and membrane transport protein; petroleum spill;
XX hydrocarbon degradation; gram positive aerobic bacterium; marker;
```

```
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
XX Corynebacterium glutamicum.
XX
XX WO200100805-A2.
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-IB000926.
XX
XX 25-JUN-1999; 99US-0141031P.
XX 08-JUL-1999; 99DE-01031454.
XX 08-JUL-1999; 99DE-01031478.
XX 08-JUL-1999; 99DE-01031563.
XX 09-JUL-1999; 99DE-01032122.
XX 09-JUL-1999; 99DE-01032124.
XX 09-JUL-1999; 99DE-01032125.
XX 09-JUL-1999; 99DE-01032128.
XX 09-JUL-1999; 99DE-01032180.
XX 09-JUL-1999; 99DE-01032182.
XX 09-JUL-1999; 99DE-01032190.
XX 09-JUL-1999; 99DE-01032191.
XX 09-JUL-1999; 99DE-01032209.
XX 09-JUL-1999; 99DE-01032212.
XX 09-JUL-1999; 99DE-01032227.
XX 09-JUL-1999; 99DE-01032228.
XX 09-JUL-1999; 99DE-01032229.
XX 09-JUL-1999; 99DE-01032230.
XX 14-JUL-1999; 99DE-01032927.
XX 14-JUL-1999; 99DE-01033005.
XX 14-JUL-1999; 99DE-01033006.
XX 27-AUG-1999; 99DE-01040764.
XX 27-AUG-1999; 99DE-01040765.
XX 27-AUG-1999; 99DE-01040766.
XX 27-AUG-1999; 99DE-01040810.
XX 27-AUG-1999; 99DE-01040831.
XX 27-AUG-1999; 99DE-01040832.
XX 27-AUG-1999; 99DE-01040833.
XX 31-AUG-1999; 99DE-01041378.
XX 31-AUG-1999; 99DE-01041379.
XX 31-AUG-1999; 99DE-01041395.
XX 03-SEP-1999; 99DE-01042077.
XX 03-SEP-1999; 99DE-01042078.
XX 03-SEP-1999; 99DE-01042079.
XX 03-SEP-1999; 99DE-01042088.
XX (BADI ) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-071486/08.
XX N-PSDB; AAF67898.
XX
Corynebacterium glutamicum nucleic acids encoding membrane construction
and membrane transport proteins or their portions, useful for typing or
identifying C. glutamicum or related bacteria, and as markers for
transformation.
XX
XX Claim 20; Page 612-614; 1119pp; English.
XX
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
construction and membrane transport (MCT) proteins given in AAB76510 to
AAB76847. The MCT nucleic acids and proteins are useful in the
identification of microorganisms which can be used to produce fine
chemicals, for modulating fine chemical production in C. glutamicum or
related bacteria (e.g. Brevibacterium lactofermentum), the typing or
identification of C. glutamicum or related bacteria, as reference points
for mapping C. glutamicum genome, and as markers for transformation.
XX
XX AAF68082 and AAF68082 represent sequencing primers which are used in an
example from the present invention
XX
XX Sequence 713 AA;
XX
```

Query Match 7.0%; Score 118; DB 4; Length 713;  
 Best Local Similarity 25.5%; Pred. No. 0.026;  
 Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLALAGLPGDRWDSTTPRLGVQKIMDWSGE---HWAK-----P 75  
 DB 172 KFNIRCVRTPELSALGPRDQGVTI---LGVKFAQQTGPTTWEKDDRRKLGDRST 227

QY 76 YATGSRDFRKTLRWQVNGFAVLN-ADNLNIATNSQLNE---YCLSDALQALRAYG 130  
 DB 228 YRLGSTNDKAVETLRETQKAVVQVQADNRITAAANRAELRELERQYQASQETILKVSQAQI 287

QY 131 TEGFESLVVFLD-----EASKAVKARAEALQAAAMISVDLPQGEFF 171  
 DB 288 DYESADAATAELDRLEELNNTPEATELSARHEAAKQTLARVSDLLVAAQSEETVASMML 347

QY 172 LLSAPAGNPLKKMVEEFVPRFAPRSTVLYLGDTRGKHSLFEREIFEVLGLTFD 226  
 DB 348 KRAETELKRLSLPVAEVSEETAREVEKLFANTRRVHA---ANVDEQITIALRED 399

RESULT 3  
 AAG92827  
 ID AAG92827 standard; protein; 1111 AA.  
 XX  
 AC AAG92827;  
 DT 26-SEP-2001 (first entry)  
 DE  
 DE C glutamicum protein fragment SEQ ID NO: 6581.  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PF 18-DEC-2000; 2000EP-00127688.  
 XX  
 PR 16-DEC-1999; 99JP-00377484.  
 PR 07-APR-2000; 2000JP-00159162.  
 PR 03-AUG-2000; 2000JP-00280986.  
 XX  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 PA  
 PI Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;  
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 XX  
 DR WPI; 2001-376931/40.  
 DR N-PSDB; AAH68046.  
 XX

Novel polynucleotides derived from Corynebacterium bacteria, for identifying mutation point of a gene, measuring expression of a gene, analyzing expression profile or pattern of a gene and identifying homologous gene.

Claim 17; SEQ ID NO 6581; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a mutant of corynebacterium bacterium, measuring expression amount and analyzing the expression profile or expression pattern of a gene derived from Corynebacterium bacterium, and identifying a homologue of a gene derived from corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a protein described in the exemplification of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office

XX  
 SQ Sequence 1111 AA;  
 Query Match 7.0%; Score 118; DB 4; Length 1111;  
 Best Local Similarity 25.5%; Pred. No. 0.052;  
 Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLALAGLPGDRWDSTTPRLGVQKIMDWSGE---HWAK-----P 75  
 DB 570 KFNIRCVRTPELSALGPRDQGVTI---LGVKFAQQTGPTTWEKDDRRKLGDRST 625

QY 76 YATGSRDFRKTLRWQVNGFAVLN-ADNLNIATNSQLNE---YCLSDALQALRAYG 130  
 DB 626 YRLGSTNDKAVETLRETQKAVVQVQADNRITAAANRAELRELERQYQASQETILKVSQAQI 685

QY 131 TEGFESLVVFLD-----EASKAVKARAEALQAAAMISVDLPQGEFF 171  
 DB 686 DYESADAATAELDRLEELNNTPEATELSARHEAAKQTLARVSDLLVAAQSEETVASMML 745

QY 172 LLSAPAGNPLKKMVEEFVPRFAPRSTVLYLGDTRGKHSLFEREIFEVLGLTFD 226  
 DB 746 KRAETELKRLSLPVAEVSEETAREVEKLFANTRRVHA---ANVDEQITIALRED 797

RESULT 4  
 ABM69452  
 ID ABM69452 standard; protein; 135 AA.  
 XX  
 AC ABM69452;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photorhabdus luminescens protein sequence #2549.  
 XX  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 XX (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 DR WPI; 2003-148459/14.  
 XX  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 XX Claim 2; SEQ ID NO 2549; 1205pp; French.  
 PS  
 CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.

XX

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:

- (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated

polypeptide or its fragment whose expression is inhibited by the  
antigenic nucleic acid; (4) an antibody capable of specifically binding  
the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
proliferation or the activity of a gene in an operon required for  
proliferation; (7) identifying a compound that influences the activity of  
the gene product or that has an activity against a biological pathway  
required for proliferation, or that inhibits cellular proliferation; (8)  
identifying a gene required for cellular proliferation or the biological  
pathway in which a proliferation-required gene or its gene product lies  
or a gene on which the test compound that inhibits proliferation of an  
organism acts; (9) manufacturing an antibiotic; (10) profiling a  
compound's activity; (11) a culture comprising strains in which the gene  
product is overexpressed or underexpressed; (12) determining the extent  
to which each of the strains is present in a culture or collection of  
strains; or (13) identifying the target of a compound that inhibits the  
proliferation of an organism. The antisense nucleic acids are useful for  
identifying proteins or screening for homologous nucleic acids required  
for cellular proliferation to isolate candidate molecules for rational  
drug discovery programs, or for screening homologous nucleic acids  
required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
*K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
the target prokaryotic essential genes. Note: The sequence data for this  
patent did not form part of the printed specification, but was obtained  
in electronic format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

Sequence 1120 AA;

Query Match 6.1%; Score 103.5; DB 6; Length 1120;  
Best Local Similarity 22.7%; Pred. No. 1.5;  
Matches 61; Conservative 41; Mismatches 82; Indels 85; Gaps 14;  
QY 82 EDPRKKTLEWDVNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
DB 573 KEIEYVWKKNDNAIVNCNMENIDPVGHTGDSIVVAPSQTLSVDEYQMLRDVSLKVR 632  
QY 128 AXGTGEG-----FESLSVFLDEASKAKARAEALQ-----AAMISVDLPGG 168  
DB 633 ALGIEGCVQVQLALDPHLSNYYIIIEVNPV--SHSSALASKATGYPIAKLAIAVGLTLD 691  
QY 169 EEFLLSP-----AGQNPLLLKMWEEFVPRF-----APRSTVLYLGDTR 206  
DB 692 E--MLNPITGTSYAAPEPTLDYVISK-IPRFPDKPEKGERELGTQMKATGEVMAIGRT- 747  
QY 207 GKHSULFEREIEFEVLGLTGD-----PHGRMPDL-----ILHDEVGRWMLFMEAVKSKG 254  
DB 748 -----YEESLLKAIRSLFVGVHGLPNGESYELDYIKERIGHQDDERLFFIGEAIR--- 799  
QY 255 PFDEERHRSLOELFVTPSAGLIFVNCFN 283  
DB 800 -----RGTSLEELHNTKIDYFFLNKFN 823

RESULT 7

AAW57632  
ID AAW57632 standard; protein; 772 AA.  
AC AAW57632;  
XX

DT 17-OCT-2003 (revised)  
DT 27-AUG-1998 (first entry)

XX VP4 protein of human rotavirus isolate 116E.

XX VP7 gene; human rotavirus; rotavirus strain G9P11; bovine rotavirus;  
KW VP4 gene; mucosal immunity stimulation; infection.

XX Human rotavirus A.

OS US5773009-A.

PN 30-JUN-1998.

XX

PF 19-FEB-1997; 97US-00802141.  
XX  
PR 15-APR-1994; 94US-00231041.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Bhan MK, Das BK, Glass RI, Gentsch JR;  
XX  
XX WPI; 1998-386933/33.  
DR N-PSDB; AAV24477.  
DR  
XX  
PT New isolated rota-virus strain G9P11 - used for stimulating immunity  
PT against infection by virulent strains of rota-viruses.  
XX  
PS Example 3; Col 33-38; 24pp; English.  
XX  
CC This sequence is the VP4 protein of human rotavirus strain 116E. The DNA  
CC sequence can be used in the rotavirus of the invention which is an  
CC isolated rotavirus strain G9P11 comprising a bovine viral VP4 virus gene  
CC and a human VP7 virus gene. The new avirulent rotavirus strain G9P11 can  
CC be used for stimulating mucosal immunity against infection by virulent  
CC strains of rotaviruses, e.g. strain serotype G10P11. Rotaviruses are  
CC responsible for diarrhoea in infants, especially in developing countries.  
CC It can also be used for producing antigens and antibodies which can be  
CC used for diagnosis. (Updated on 17-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 772 AA;

Query Match 5.9%; Score 100; DB 2; Length 772;  
Best Local Similarity 22.6%; Pred. No. 1.8;  
Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;

QY 1 VNSSDGDIDGTASIDTARALLKRFGEFDAQYVVRSAVTLALA-----GLKPGDRWVDST 55  
DB 473 VPSNDDYQTPHANSVTVRODLER-QLDENMRREFNELSANIALSQLIDLALLP-----LDMF 527  
QY 56 TPLRGVQKIMDSGEHWAKPYATGSREDPRK-----KTLRQ 91  
DB 528 SMFSGIRSTIE-----AAKNFATSVMKPKRKNLAKSVNSLTDAITDAAGSISRSTLS 582  
QY 92 -----WYDNGFAVLNADNLNATNSQLNNEYCLSDALQALRAYGTGEGF--EESLVVFL 142  
DB 583 ANSAVSVMWTDISDVIDSTDNVVTAT-----ATAAAKPRVKFEFTTEFGVSD 630  
QY 143 DEASKAVKARAEALQAAAMISVDLPGEELLSFAGQNPPLKKMWEEFVPRFAPR---STV 199  
DB 631 DISAAVVKTKMKNLV-----VD-----EEML-----PQIITEASEKFIENRAYRLIDGDK 676  
QY 200 LYLGDTRGKHSULFEREIEFEVLGLTDFPHGRMPDLILHDEVGRWMLFMEAVKS-KGPFDE 258  
DB 677 VYEVTTGKYFAYLTETFEVM---FDAE-RAELVTYSFVISAIIIDFTIKNLNDNYGI 732  
QY 259 ERHRSIQELFVTPSAGLIFVNCFNRESMRQWLPEL 294  
DB 733 TREQALNMLRSDPKVLRSPFN--QNNPIKNRIEQL 766

RESULT 8

ABU43956  
ID ABU43956 standard; protein; 1057 AA.  
XX  
AC ABU43956;  
XX

DT 19-JUN-2003 (first entry)

XX Protein encoded by Prokaryotic essential gene #29483.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

OS Staphylococcus haemolyticus.

PN WO200277183-A2.

XX

PD 03-OCT-2002.

XX 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2001; 2001US-00915242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI; 2003-029926/02.

DR N-PSDB; ACA47826.

XX New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

XX Claim 25; SEQ ID NO 71880; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1057 AA;

Query March 5.9%; Score 99.5; DB 6; Length 1057; Best Local Similarity 21.9%; Pred. No. 3.4; Matches 59; Conservative 42; Mismatches 83; Indels 85; Gaps 14;

QY 82 EDRKKTUQWNGFAVINDNLN-----IATNSQLN--EY-CLSDALQALR 127

DB 214 KEIEVNRKNDNAIVVCMENIDPVGIHTGDSIVAPSTQSLDSVEYQMLRVSCLKVIR 273

QY 128 AYGTEG-----FEESLVVFLDEASKAVKARALQ-----AMISVDLPGG 168

DB 274 ALGIEGCGNQLALDPHFNVIYEVNPRV-SRSSALASKATGYPIAKLAIAVGLTLD 332

QY 169 EEFILSP-----AGQNPLKKWVEEFVPRF-----APRSTVLVLDTR 206

DB 333 E--MLNPVTGTSYAAFEETLDYVISK-IFRPFPDFKEKGERLGTQMKATGEVMAIGT- 388

QY 207 GKHSLFEREFEFVGLTFTD-----PHGRMPDL-----ILHDEVRGWLFLEAVKSKG 254

DB 389 -----YEESLLKAIRSLGVHHLGNGESFDLDYIKERISHQDDRLRPFGEAIR--- 440

QY 255 PFDEERHRSQELFVTPSAGLIFVNCFEN 283

DB 441 -----RGTTLEEIHNTQIDYFFLNKFN 464

RESULT 9

AAU36347

ID AAU36347 standard; protein; 289 AA.

XX AC AAU36347;

XX 14-FEB-2002 (first entry)

XX Pseudomonas aeruginosa cellular proliferation protein #337.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

XX antibacterial; drug design.

XX Pseudomonas aeruginosa.

OS WO200170955-A2.

PN 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ; Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS54206.

XX New polynucleotides for the identification and development of antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 11940; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 289 AA;

```
Query Match      5.8%; Score 97.5; DB 4; Length 289;
Best Local Similarity 24.0%; Pred. No. 0.72;
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNLNATNSQLNEYCLSDLEALQALRAYGTGFEESLVVFLDEASKAVKARA 153
Db 69 DNKAIVI-----IEVNSQTDFLALQDD-----FKGFVAESLEKAFNEKLTDAAPLVEARE 118

QY 154 EALQAAM-----ISVDLPGBGEFLLSPAGQNPLLK 183
Db 119 EARLALVAKTGENVNIRLRTVEGDVVGVYLVHGRIGVVVNLKGG-----NPELA 168

QY 184 KMVEEFV-----PRAPRSTVLYLGDTGRKHSLEPERIFEVLGLTFDPHGRMPDLILHDE 239
Db 169 KDIAHVAASNPOFLSASEVSEEAIAK-----EKEIF---LALNADKIAGKPNIVENM 219

QY 240 VRGWL--FLMEAVKSGPFDEERHSLOELFVTPSAGLIFVNCFFENRESMRQWLPDLAW 297
Db 220 VKGRISKFLAEASLVQPFVKNPEVKVDLAKQAGABIVSFVRYEGEGIEKAEVDFAAE 279

QY 298 TEAWVA 303
Db 280 VAAQVA 285

RESULT 10
ABU38597
ID ABU38597 standard; protein; 289 AA.
AC ABU38597;
XX
DT 19-JUN-2003 (first entry)
DE Protein encoded by Prokaryotic essential gene #24124.
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
PI Wall D, Irawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA42467.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 66521; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
```

```
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 289 AA;
```

```
Query Match      5.8%; Score 97.5; DB 6; Length 289;
Best Local Similarity 24.0%; Pred. No. 0.72;
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNLNATNSQLNEYCLSDLEALQALRAYGTGFEESLVVFLDEASKAVKARA 153
Db 69 DNKAIVI-----IEVNSQTDFLALQDD-----FKGFVAESLEKAFNEKLTDAAPLVEARE 118

QY 154 EALQAAM-----ISVDLPGBGEFLLSPAGQNPLLK 183
Db 119 EARLALVAKTGENVNIRLRTVEGDVVGVYLVHGRIGVVVNLKGG-----NPELA 168

QY 184 KMVEEFV-----PRAPRSTVLYLGDTGRKHSLEPERIFEVLGLTFDPHGRMPDLILHDE 239
Db 169 KDIAHVAASNPOFLSASEVSEEAIAK-----EKEIF---LALNADKIAGKPNIVENM 219

QY 240 VRGWL--FLMEAVKSGPFDEERHSLOELFVTPSAGLIFVNCFFENRESMRQWLPDLAW 297
Db 220 VKGRISKFLAEASLVQPFVKNPEVKVDLAKQAGABIVSFVRYEGEGIEKAEVDFAAE 279

QY 298 TEAWVA 303
Db 280 VAAQVA 285
```

```
RESULT 11
AAY05897
ID AAY05897 standard; protein; 513 AA.
XX
AC AAY05897;
XX
XX 02-AUG-1999 (first entry)
XX
DT Vicia sativa omega-myristic acid hydroxylase CYP94A2.
DE Fatty acid omega-hydroxylase; omega-myristic acid hydroxylase;
XX cytochrome P450; transgenic plant; lipid; hydroxylation; epoxidation;
KW oilseed; vegetable oil; crop protection; omega-hydroxy acid; CYP94A2.
XX
OS Vicia sativa.
XX
FH Key Location/Qualifiers
FT Domain 319..329
FT /note= "haem-binding domain, corresponds to motif claimed
XX in Claim 2"
XX
PN WO9918224-A1.
```

```

XX 15-APR-1999.
XX
XX
XX 06-OCT-1998; 98WO-IB001716.
XX
XX 06-OCT-1997; 97US-0060960P.
XX
XX (CNRS ) CENT NAT RECH SCI.
XX
XX Tijet N, Pinot F, Benveniste I, Le Bouquin R, Helvig C, Batard Y,
XX Cabello-Huatao F, Werck-Reichhart D, Salaun J, Durst F;
XX
XX WPI; 1999-264030/22.
XX
XX N-PSDB; AAX58401.
XX
XX Nucleic acid encoding plant fatty acid hydroxylases.
XX
XX Example 2; Fig 7; 157pp; English.
XX
XX The present sequence represents Vicia sativa CYP94A2 omega-myristic acid
XX hydroxylase, a microsomal cytochrome P450-dependent hydroxylase which
XX catalyses the transformation of myristic acid into 14-
XX hydroxytetradecanoic acid (i.e. terminal methyl hydroxylation). Low
XX levels of transformation of lauric and palmitic acids into the
XX corresponding omega-hydroxy fatty acids are also observed. The invention
XX provides isolated nucleic acids (see AAX58400-06) encoding plant fatty
XX acid hydroxylases (see AAY05896-902). Also claimed are host cells,
XX transgenic plants and compositions consisting of the plant fatty acid
XX hydroxylase, a process for isolating additional fatty acid hydroxylase
XX genes from a plant, and a process of altering fatty acid composition in a
XX plant by expressing the plant fatty acid hydroxylase in a transgenic
XX plant, and hydroxylating or epoxidating a fatty acid substrate in the
XX plant. Manipulating the hydroxylated fatty acid content of plants will
XX modify resistance to drought and attack by insects and other pests. The
XX transgenic plants may also be used as sources of hydroxylated and
XX epoxidized fatty acids useful in the manufacture of e.g. lubricants, anti
XX -slip agents, plasticisers, coating agents, detergents and surfactants
XX
XX Sequence 513 AA;
XX
XX Query Match 5.7%; Score 96.5; DB 2; Length 513;
XX Best Local Similarity 22.6%; Pred. No. 2.2;
XX Matches 72; Conservative 44; Mismatches 112; Indels 91; Gaps 15;
XX
XX QY 64 IMDWGEHWKPYATGSRDPRKTKLROWVNGFAVLNADN-----LNATNSQLN-----114
XX Db 126 IFNADGESWKTQRQISSEH-FNTRSLRKFVETWVDVLSDRPLVPVLSQANSQTTLDFQD 184
XX
XX QY 115 -----EYC-----LSDEALQAL-----RAYG 130
XX Db 185 ILQRLTFDNCIARGYDPEVLLPSLPFAKAFDESSQSLSIERLNALIPLLWKVKRFL 244
XX
XX QY 131 TEGFEESLVFLDE-----ASKAVK-----ARAEALQAAIMI SVDLPGGEEFLSPAGQNP 181
XX Db 245 NIGVERQLKEAAVAVRGLATKIVKNKKELKEALQSESVDDL--SRFLSSGHSDES 302
XX
XX QY 182 LKKWVEEFVPPAPRST-----VLYLGDTRGKHSLEPERIFEVLGLTFDPHGMEDLLH 237
XX Db 303 VTDWVISITII--LAGRDTTSAALTWFFLLSKSHSHVENEILKEITGKS-----ETGY 352
XX
XX QY 238 DEVRGWLF-----LMEAVSKSGPFDEERHRSIQELFVTPSAGI-----IFVNCFFNRBSM 287
XX Db 353 DEVKDWYTHAALCESMRLLPPLPVDTKVAVHD-DVLPDGLTKKGRVTVYHIYAMGRSE 411
XX
XX QY 288 ROWLPALA-WETEAWVAED 305
XX Db 412 KIWGPDWAEFRPRERWLSRD 430
XX
XX RESULT 12
XX AAU33575
XX ID AAU33575 standard; protein; 797 AA.
XX

```

```

AC AAU33575;
XX
XX 14-FEB-2002 (first entry)
XX
XX Pseudomonas aeruginosa cellular proliferation protein #19.
XX
XX Antisense; prokaryotic cellular proliferation protein; antibiotic;
XX antibacterial; drug design.
XX
XX Pseudomonas aeruginosa.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009180.
XX
XX 21-MAR-2000; 2000US-0191078P.
XX
XX 23-MAY-2000; 2000US-0206848P.
XX
XX 26-MAY-2000; 2000US-0207727P.
XX
XX 23-OCT-2000; 2000US-0242578P.
XX
XX 27-NOV-2000; 2000US-0253625P.
XX
XX 22-DEC-2000; 2000US-0257931P.
XX
XX 16-FEB-2001; 2001US-0269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX N-PSDB; AAS51434.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids.
XX
XX Example 3; SEQ ID NO 5071; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the genes,
XX themselves and the encoded proteins. The prokaryotes used are Escherichia
XX coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
XX Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
XX useful for the identification of potential new targets for antibiotic
XX development. The antisense nucleic acids can also be used to identify
XX proteins used in proliferation, to express these proteins, and to obtain
XX antibodies capable of binding to the expressed proteins. The proteins can
XX be used to screen compounds in rational drug discovery programmes. The
XX antisense nucleic acid sequence is also useful to screen for homologous
XX nucleic acids which are required for cell proliferation in a wide variety
XX of organisms. The present sequence represents an essential prokaryotic
XX cellular proliferation protein. Note: The sequence data for this patent
XX did not form part of the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 797 AA;
XX
XX Query Match 5.7%; Score 95.5; DB 4; Length 797;
XX Best Local Similarity 24.7%; Pred. No. 5.5;
XX Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;
XX
XX QY 10 TVASIDTARALLKRGFDQAQ-RYNVRSAVTLLA-----LAGLKPG---DRWVDST 55
XX Db 325 TQFSIDNSTLGLWVNDWSRVYANRAEQMLGHADGQLVDRPLADFEPLGDMRWLNL- 383
XX
XX QY 56 TPRLGQKIMDWSGEHWKPYATGSR-----DPRKTKLROWVNGFAVLNADNLTNSQ 112
XX Db 384 -----WRR--ARNSEEGFLSPFETCLR-----AD-----405
XX
XX 113 LNEYCLSDALQALRAYGTGTEGFEESLVVFLDEASKAVKARAEALQ--AAM--ISVDLP 167
XX

```

Db 406 -GSMLEPADVLSFLR-FGT---SEYLVVFLSDVTTERRR- EALQSEARMKGIASNVPG 459  
 QY 168 GEEFLLSPAGONPLKKWVEFV-PRFAPRSTVLYLGDTRGKHSLF---EREIFEVLGL 223  
 Db 460 -----MVFLEPRPRAGATSDPAYISE--GSEALVGYSAKELIESGRGI 500  
 QY 224 TFDPHGRMPDILLHDEVGRGLFMEAVKSGPPDEERHRSIQELFTVPSAGLIFVN---- 279  
 Db 501 RGLVH---PD---DRERYWSSQMAAL-----DENRDWHQGRILTRQGLRWADIKAS 547  
 QY 280 --CFENRESRMQWLPELAWETEAW 301  
 Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 13  
 ABU15621  
 ID ABU15621 standard; protein; 797 AA.  
 XX AC ABU15621;  
 XX DT 19-JUN-2003 (first entry)  
 XX DE Protein encoded by Prokaryotic essential gene #1148.  
 XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX OS Pseudomonas aeruginosa.  
 XX PN WO200277183-A2.  
 XX PD 03-OCT-2002.  
 XX PF 21-MAR-2002; 2002WO-US009107.  
 XX PR 21-MAR-2001; 2001US-00815242.  
 XX PR 06-SEP-2001; 2001US-00948993.  
 XX PR 25-OCT-2001; 2001US-0342923P.  
 XX PR 08-FEB-2002; 2002US-00072851.  
 XX PR 06-MAR-2002; 2002US-0362699P.  
 XX PA (ELIT-) ELTRA PHARM INC.  
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 XX WPI: 2003-029926/02.  
 XX N-PSDB; ACA19491.  
 XX PT New antisense nucleic acids, useful for identifying proteins or screening  
 XX PT for homologous nucleic acids required for cellular proliferation to  
 XX PT isolate candidate molecules for rational drug discovery programs.  
 XX PS Claim 25; SEQ ID NO 43545; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene

CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 797 AA;

Query Match 5.7%; Score 95.5; DB 6; Length 797;  
 Best Local Similarity 24.7%; Pred. No. 5.5;  
 Matches 80; Conservative 36; Mismatches 91; Indels 117; Gaps 22;  
 QY 10 TVASIDTARALLKRFQFDAQ-RYNVRSVNTLLA-----LAGLKPQ---DRWVDST 55  
 Db 325 TQFSIDNSTLIGILVNMVDSRVYANRAAEQMLGHADGQLVDRPLADPEGLDMRWLNL- 383  
 QY 56 TPLGVQKIMDWSGEHWAKPYATGSR---DREKKTILROWVDNGFAVLNADNLNIATNSQ 112  
 Db 384 -----WRR--ARNSEEGPLSFETCLR-----AD----- 405  
 QY 113 LNEYCLSDALQALRAYGTGFEESLVVFLDEASKAVKARAEALQ---AAM--ISVDLPQ 167  
 Db 406 -GSMLEPADVLSFLR-FGT---SEYLVVFLSDVTTERRR- EALQSEARMKGIASNVPG 459  
 QY 168 GEEFLLSPAGONPLKKWVEFV-PRFAPRSTVLYLGDTRGKHSLF---EREIFEVLGL 223  
 Db 460 -----MVFLEPRPRAGATSDPAYISE--GSEALVGYSAKELIESGRGI 500  
 QY 224 TFDPHGRMPDILLHDEVGRGLFMEAVKSGPPDEERHRSIQELFTVPSAGLIFVN---- 279  
 Db 501 RGLVH---PD---DRERYWSSQMAAL-----DENRDWHQGRILTRQGLRWADIKAS 547  
 QY 280 --CFENRESRMQWLPELAWETEAW 301  
 Db 548 ARCFEDGRAV-----WDGVVW 563

RESULT 14  
 ABU18981  
 ID ABU18981 standard; protein; 1057 AA.  
 XX AC ABU18981;  
 XX DT 06-MAR-2003 (first entry)  
 XX DE Pathogen specific antigen related staphylococcal protein SEQ ID No 152.  
 XX KW Antibacterial; virucide; fungicide; protozoacide; cytostatic; anti-HIV;  
 XX KW hyperimmune; serum-reactive; antigen; pathogen; tumour; allergen;  
 XX KW auto-immunity; vaccine; staphylococcal infection; antibody; cancer;  
 XX KW autoimmune disease; HIV; hepatitis.  
 XX OS Staphylococcus sp.  
 XX PN WO200259148-A2.  
 XX PD 01-AUG-2002.  
 XX PF 21-JAN-2002; 2002WO-EP000546.  
 XX PR 26-JAN-2001; 2001AT-00000130.  
 XX PA (CIST-) CISTEM BIOTECHNOLOGIES GMBH.  
 XX PI Meinke A, Nagy E, Von Ahsen U, Klade C, Henics T, Zauner W;

PI Minh DB, Vytvytska O, Etz H, Dryla A, Weichhart T, Hafner M;  
PI Tempelmaier B;  
XX WPI; 2003-075410/07.  
XX Identifying, isolating and producing hyperimmune serum-reactive antigens  
XX from a pathogen, for preparing vaccine or medicament for treating or  
XX preventing e.g. staphylococcal infections, comprises providing antibody  
XX preparation.  
XX Claim 21; Page 173; 252pp; English.  
XX The invention relates to a novel method for identifying, isolating and  
XX producing hyperimmune serum-reactive antigens from a pathogen, tumour,  
XX allergen, a tissue or host prone to auto-immunity, where the antigens are  
XX used in a vaccine, comprises providing antibody preparation from a plasma  
XX pool of a type of animal, or individual sera with antibodies against the  
XX specific pathogen, tumour, allergen, tissue or host prone to auto-  
XX immunity. The hyperimmune serum-reactive antigens comprising any of the  
XX 62 sequences of 53-2261 amino acids fully defined in the specification,  
XX or their hyperimmune fragments are useful for the manufacture of a  
XX pharmaceutical preparation, particularly a vaccine against staphylococcal  
XX infections or colonisation against *S. aureus* or *S. epidermidis*. The  
XX preparation of antibodies is useful for the manufacture of a medicament  
XX for treating or preventing staphylococcal infections or colonisation  
XX against *S. aureus* or *S. epidermidis*. The antibody preparations may also  
XX be used for diagnostic and imaging purposes. Other conditions that can be  
XX treated include cancer, autoimmune diseases or infections caused by viral  
XX (e.g. HIV, hepatitis A, B or C), fungal or protozoan pathogens. This  
XX sequence represents a staphylococcal protein relating to the method for  
XX identifying and producing pathogen specific antigens of the invention  
XX  
XX Sequence 1057 AA;  
Query Match 5.7%; Score 95.5; DB 6; Length 1057;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;  
QY 82 EDFRKKTLROWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
Db 214 KEIEYVNRKNDNAIVVCNMENIDPVGHTGDSIVAPQSOTLSDEYQMLRDSVKVIR 273  
QY 128 AYGTGG-----FEESLVWFLDEASKAVKARAEALQ-----AAMISVDLPFG 168  
Db 274 ALGIEGGCNVQLALDPHSPDYIIIEVNPV-SRSSALASKATGYPIAKLAIAVGLTLD 332  
QY 169 EEFLSP-----AGQNPLKKWVEFVPRF-----APRSTVLYLGDT 206  
Db 333 E--MLNPITGTSYAAFEPTLDYVSK-IPRFPFKPKGERELGTQMKATGEVMAIGRT- 388  
QY 207 GKHSIFEREIPEVGLTFD-----PHGRMPDL-----ILHDEVRGWLFLEAVKSG 254  
Db 389 -----YEESLLKAIRSLGYVHHLGPNGESFDLDYIKERISHQDDERLFFGEAIR- 440  
QY 255 PFDEERHRSLOSLFTVPSAGLIFVNCFNRESMROWLPE 293  
Db 441 -----RGYTLSEIHNTQIDYFLFKFQNIIDIEHQKE 474  
RESULT 15  
ABU42381  
ID ABU42381 standard; protein; 1057 AA.  
XX  
AC ABU42381;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
XX Protein encoded by Prokaryotic essential gene #27908.  
XX Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX Staphylococcus aureus.  
XX

PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-USO09107.  
XX  
XX 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
XX (ELIT-) ELITRA PHARM INC.  
XX  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA46251.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX  
XX Claim 25; SEQ ID NO 70305; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 1057 AA;  
Query Match 5.7%; Score 95.5; DB 6; Length 1057;  
Best Local Similarity 21.5%; Pred. No. 8.5;  
Matches 60; Conservative 44; Mismatches 90; Indels 85; Gaps 14;  
QY 82 EDFRKKTLROWVDNGFAVLNADNLN-----IATNSQLN--EY-CLSDALQALR 127  
Db 214 KEIEYVNRKNDNAIVVCNMENIDPVGHTGDSIVAPQSOTLSDEYQMLRDSVKVIR 273  
QY 128 AYGTGG-----FEESLVWFLDEASKAVKARAEALQ-----AAMISVDLPFG 168  
Db 274 ALGIEGGCNVQLALDPHSPDYIIIEVNPV-SRSSALASKATGYPIAKLAIAVGLTLD 332  
QY 169 EEFLSP-----AGQNPLKKWVEFVPRF-----APRSTVLYLGDT 206

Db 333 E--MLNPIITGTSYAAFEFTLDYVISK-IPRPPDKFKGERELGTOMKATGEVMAIGRT- 388  
QY 207 GKHSLPEREIPEEVLGTED-----PHGRMPDL-----ILHDEVRCWLFLEAVKSKG 254  
Db 389 -----YEESLLKAIRSLEYGVHHLGLFNGESFDLDYKERISHODDERLFFIGEAIR--- 440  
QY 255 PFDEERHRSLOELFVTPESAGLIFVNCFFENRESMRQWLPE 293  
Db 441 -----RGTTLEEIHNMTCIDYFFLHKFQNIIDIEHQLKE 474

Search completed: October 1, 2004, 16:13:00  
Job time : 47.3214 secs

Blank

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 1, 2004, 16:09:01 ; Search time 13.1755 Seconds  
(without alignments)  
1265.619 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIQGVASIDTARAL.....EDPDHLHLNGSRFLQGYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/prodata/2/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pcp.\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	6.1	1037	US-09-134-001C-4794	Sequence 4794, Ap
2	100	5.9	772	US-08-802-141-4	Sequence 4, Appli
3	97.5	5.8	310	US-09-252-991A-21111	Sequence 21111, A
4	96	5.7	511	US-08-676-444-42	Sequence 42, Appl
5	95.5	5.7	835	US-09-252-991A-17364	Sequence 17364, A
6	93	5.5	498	US-08-470-202-59	Sequence 59, Appl
7	93	5.5	498	US-08-471-770-59	Sequence 59, Appl
8	93	5.5	498	US-08-468-059-59	Sequence 59, Appl
9	93	5.5	498	US-09-109-916-59	Sequence 59, Appl
10	93	5.5	498	US-09-886-156-59	Sequence 59, Appl
11	93	5.5	498	US-09-886-149-59	Sequence 59, Appl
12	93	5.5	498	US-09-886-150-59	Sequence 59, Appl
13	93	5.5	498	US-09-886-159-59	Sequence 59, Appl
14	92	5.4	171	US-09-540-236-2189	Sequence 59, Appl
15	91.5	5.4	1360	US-09-788-657-22	Sequence 22, Appl
16	91	5.4	506	US-09-299-662-4	Sequence 4, Appli
17	91	5.4	524	US-09-299-662-1	Sequence 1, Appli
18	90.5	5.4	503	US-09-408-020-66	Sequence 66, Appl
19	90	5.3	372	US-09-973-963-3	Sequence 3, Appli
20	88.5	5.2	501	US-09-489-039A-12663	Sequence 4, Appli
21	86.5	5.1	561	US-09-252-991A-16726	Sequence 16726, A
22	86	5.1	466	US-09-252-991A-32781	Sequence 32781, A
23	86	5.1	3472	US-09-408-020-4	Sequence 4, Appli
24	84	5.0	450	US-09-134-001C-4811	Sequence 4811, Ap
25	84	5.0	822	US-09-886-319A-64	Sequence 64, Appl
26	83.5	4.9	877	US-08-208-036-14	Sequence 14, Appl
27	83.5	4.9	877	US-08-428-823-14	Sequence 14, Appl

28	83	4.9	295	4	US-09-252-991A-33101	Sequence 33101, A
29	83	4.9	348	3	US-08-855-910-8	Sequence 8, Appli
30	83	4.9	498	1	US-08-470-202-60	Sequence 60, Appl
31	83	4.9	498	1	US-08-471-770-60	Sequence 60, Appl
32	83	4.9	498	2	US-08-468-059-60	Sequence 60, Appl
33	83	4.9	498	3	US-09-109-916-60	Sequence 60, Appl
34	83	4.9	498	4	US-09-886-156-60	Sequence 60, Appl
35	83	4.9	498	4	US-09-886-149-60	Sequence 60, Appl
36	83	4.9	498	4	US-09-886-150-60	Sequence 60, Appl
37	83	4.9	498	4	US-09-886-159-60	Sequence 60, Appl
38	82.5	4.9	876	2	US-08-633-476-2	Sequence 2, Appli
39	82	4.9	223	4	US-09-489-039A-12066	Sequence 12066, A
40	82	4.9	893	4	US-09-514-302-4	Sequence 4, Appli
41	82	4.9	1938	4	US-09-514-302-2	Sequence 2, Appli
42	81.5	4.8	320	4	US-09-134-000C-5021	Sequence 5021, Ap
43	81.5	4.8	331	1	US-08-134-570-12	Sequence 12, Appl
44	81.5	4.8	479	1	US-08-665-220-2	Sequence 2, Appli
45	81.5	4.8	479	3	US-09-291-692-2	Sequence 2, Appli

## ALIGNMENTS

## RESULT 1

US-09-134-001C-4794  
; Sequence 4794, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4794  
; LENGTH: 1037  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-4794

Query Match 6.1%; Score 103.5; DB 4; Length 1037;  
Best Local Similarity 22.7%; Pred. No. 0.1;  
Matches 61; Conservative 41; Mismatches 82; Indels 85; Gaps 14;  
QY 82 EDFRKKTLROWVDNGFAVLNADNLN-----IATNSOLN--EY-CLSEALQALR 127  
Db 224 KEIEYEVNRDKNDNAIVVNCNMENIDFVGHTGDSIVVAPSPQTLSDVEYQMLRDSVLKVR 283  
QY 128 AYTGTG-----FEESLWFLDEASKAVKARALQ-----AAMISVDLPQ 168  
Db 284 ALGIEGGCNVQLALPHSLNLYIIVNPRV-SRSSALASKATGYPIAKIAVGLTLD 342  
QY 169 EFLFISP-----ACQNPLKXWVEEFVPRF-----APRSTVLYLGDT 206  
Db 343 E-MLNPITGTSYAAFEPTLDYISK-IPRFPDFKFEGERELGTQMKATGEVMAIGRT- 398  
QY 207 GKHSFUREIRIEFEVULGTFD-----PHGRMPDL-----ILHDEVRGWLFLMEAVKSG 254  
Db 399 -----YEESLLKAIKRSLEYGVHHLGLPNGESVELDYIKERIGHQDDERLFFTGEAIR--- 450  
QY 255 PFDEERHESLOELFVTPSAGLIFVNCEN 283  
Db 451 -----RGTSLEELHNMTKIDYFFLNKFN 474

## RESULT 2

US-08-802-141-4  
; Sequence 4, Application US/08802141

```

; Patent No. 5773009
; GENERAL INFORMATION:
; APPLICANT: GLASS, ROGER I.
; APPLICANT: GENTSCH, JOHN R.
; APPLICANT: BHAN, M. K.
; APPLICANT: DAS, BIMAL K.
; TITLE OF INVENTION: ROTAVIRUS STRAIN AND RELATED
; COMPOSITIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/802,141
; FILING DATE: 19-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/231,041
; FILING DATE: 15-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.609
; TELEPHONE: 404/688-0770
; TELEFAX: 688-9880
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 772 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-802-141-4

Query Match      5.9%; Score 100; DB 1; Length 772;
Best Local Similarity 22.6%; Pred. No. 0.15;
Matches 76; Conservative 48; Mismatches 128; Indels 84; Gaps 16;

QY 1 VNSSDGDGTVASIDTARALLKRFQDAQRYNVRSAVTLLALA-----GLKPGDRWVDST 55
DB 473 VPSDDYQTPIANSVTVRODLER-QLDENRREFNELSANIALSOLIDLALLP---LDMF 527

QY 56 TPRLGVQKIMWSGHEWAKPYATGSRDPK-----KTLRQ 91
DB 528 SMFSGIRSTIE-----AAKNFATSVMKPKRKNLAKSVNSLTDAITDAAGSISRSTLRS 582

QY 92 -----WVNGFAVLNADNINIATNSQNLNEYCLSDALQALRAYCTGCF--EESLVVFL 142
DB 583 ANSAVSVWMTDIDSDVDSTDNVVTAT-----ATAAAKPRVKEFTEFNGVSFD 630

QY 143 DEASKAVKARAEALQAAMTSDVLPGGEEFLLSPAGQNPLLKQWVEEFVPRFAPR---STV 199
DB 631 DISAARVKTQMKNLV-----VD-----EEML-----PQITEASEKFIENRAYRLIDGDK 676

QY 200 LYLGDTRGRKHSIFEREIFEIVGLTDPHGRMPDLILHDEVRGWLFLMEAVKS-KGPFDE 258
DB 677 VYEVTTGKYFAYLTETFEVVM--FDAE-RFAELVTYSPVISAIIDFKTIKNLDNYGI 732

QY 259 EHRSLQELFVTPSAGLIIFVNCFENRESMRQWLP 294
DB 733 TREQALNMLRSDPKVLRSPIN--QNNPIIKNRIEQL 766

; RESULT 3

```

```

US-09-252-991A-21111
; Sequence 21111, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21111
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-21111

Query Match      5.8%; Score 97.5; DB 4; Length 310;
Best Local Similarity 24.0%; Pred. No. 0.067;
Matches 59; Conservative 32; Mismatches 90; Indels 65; Gaps 8;

QY 94 DNGFAVLNADNINIATNSQNLNEYCLSDALQALRAYCTGFEESLVVFLDEASKAVKARA 153
DB 90 DNKAAVI-----TEVNSQTDFLALQDD---FKGFVAESLEKAFNEKLTDAAPLIVEARE 139

QY 154 EALQAAM-----ISVDLPGEFEFLLSPAGQNPLLK 183
DB 140 EARLALVAKTCNGVNIIRRLTRVEGDVVGAYLHGHRIGVVNVLKGG-----NP 189

QY 184 KMVEEFV---PRFAPRSTVLYLGDTRCKHSLFEREIFEIVGLTDFDPHGRMPDLILHDE 239
DB 190 KDIAMHVAASNPOFLSASEVSEEAIAK-----EKEIF---LALNADKIACKPENIVNM 240

QY 240 VRGWL--FLMEAVKSKGPFDEERHRSLOELFVTPSAGLIIFVNCFENRESMRQWLP 297
DB 241 VKGRISKFLAEASLVVEQPFVKNPVEKVGDLAKQAGAEIVSFVRYVEGEGIEKAEVDFAE 300

QY 298 TEAWVA 303
DB 301 VAAQVA 306

; RESULT 4
US-08-676-444-42
; Sequence 42, Application US/08676444A
; Patent No. 6294325
; GENERAL INFORMATION:
; APPLICANT: Wetmur, James G.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
; TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
; FILE REFERENCE: MSM95-02
; CURRENT APPLICATION NUMBER: US/08/676,444A
; CURRENT FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 42
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Thermotoga maritima
; US-08-676-444-42

Query Match      5.7%; Score 96; DB 3; Length 511;
Best Local Similarity 20.8%; Pred. No. 0.21;
Matches 64; Conservative 35; Mismatches 99; Indels 110; Gaps 13;

QY 94 DNGFAVLNADNINIATNSQNLNEYCLSDALQALRAYCTGFEESLVVFLDEASKAVKARA 153
DB 59 DNGIGMTREAL-LAIEPYTTSKISSEEDLHRTYVGRFGEALASIVQVSRKIVTKTEK 117

```



Qy 102 ADNLNATNSQMEYCL----SDEALQALRAYTEG-FBESLVVFLDEASKAVARAEAL 156  
Db 305 AEQATQEVKNWMTETLLVQNSNPDCKQILKALGPATLEEMVACQGVGPTHKAILAE 364  
Qy 157 QAAMISVDLPGGEBEFLSPAGQNPLLKKMVEEF 189  
Db 365 AMASAOQDLKGGYTAVFMQGNPNRKGPIKCF 397

RESULT 8  
US-08-468-059-59  
; Sequence 59, Application US/08468059  
; Patent No. 5840480  
; GENERAL INFORMATION:  
; APPLICANT: Guertler, Lutz G.  
; APPLICANT: Eberle, Josef  
; APPLICANT: Brunn, Albrecht v.  
; APPLICANT: Knapp, Stefan  
; APPLICANT: Hauser, Hans-Peter  
; TITLE OF INVENTION: Retrovirus from the HIV Group and its  
; TITLE OF INVENTION: Use  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/468,059  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/132,653  
; FILING DATE: 05-OCT-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 33 646.5  
; FILING DATE: 06-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 35 718.7  
; FILING DATE: 22-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 44 541.8  
; FILING DATE: 30-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 43 18 186.4  
; FILING DATE: 01-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carol P. Einaudi  
; REGISTRATION NUMBER: 32,220  
; REFERENCE/DOCKET NUMBER: 05495-0001-02000  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-468-059-59

Query Match 5.5%; Score 93; DB 2; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;

Db 365 AMASAOQDLKGGYTAVFMQGNPNRKGPIKCF 397

US-08-471-770-59  
; Sequence 59, Application US/08471770  
; Patent No. 5770427  
; GENERAL INFORMATION:  
; APPLICANT: Guertler, Lutz G.  
; APPLICANT: Eberle, Josef  
; APPLICANT: Brunn, Albrecht v.  
; APPLICANT: Knapp, Stefan  
; APPLICANT: Hauser, Hans-Peter  
; TITLE OF INVENTION: Retrovirus from the HIV Group and its  
; TITLE OF INVENTION: Use  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
; STREET: 1300 I Street, N.W.  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,770  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/132,653  
; FILING DATE: 05-OCT-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 33 646.5  
; FILING DATE: 06-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 35 718.7  
; FILING DATE: 22-OCT-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 42 44 541.8  
; FILING DATE: 30-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: DE P 43 18 186.4  
; FILING DATE: 01-JUN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carol P. Einaudi  
; REGISTRATION NUMBER: 32,220  
; REFERENCE/DOCKET NUMBER: 05495-0001-03000  
; TELEPHONE: 202-408-4000  
; TELEFAX: 202-408-4400  
; INFORMATION FOR SEQ ID NO: 59:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
; US-08-471-770-59

Query Match 5.5%; Score 93; DB 1; Length 498;  
Best Local Similarity 26.8%; Pred. No. 0.42;  
Matches 41; Conservative 20; Mismatches 68; Indels 24; Gaps 6;

Qy 48 GD---RWVDSTTFLGVQKIMDSGEHWAKPVATGSRDFF---KKTLRQWVDNGFAVLN 101  
Db 258 GDYIRKWI-----VLGLNKNW-----RWVSFVSLIDIQGKPEPFRDYDRFYKTLR 304





Search completed: October 1, 2004, 16:13:58  
Job time : 15.1755 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: October 1, 2004, 16:13:07 ; Search time 69.0717 Seconds  
(without alignments)  
1504.830 Million cell updates/sec

Title: US-10-668-047-4

Perfect score: 1689

Sequence: 1 VNSSDGIDGTWASIDTARAL.....EDPDHLHLNGSRFLGPYER 323

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1351062 seqs, 321799191 residues

Total number of hits satisfying chosen parameters: 1351062

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	7.0	708	12	US-10-627-476-314
2	118	7.0	713	12	US-10-627-476-314
3	118	7.0	1111	9	US-09-738-626-6581
4	105.5	6.2	229	15	US-10-211-689-34
5	105.5	6.2	236	15	US-10-211-689-34
6	103.5	6.1	1120	12	US-10-282-122A-70903
7	103	6.1	194	15	US-10-211-689-30
8	101	6.0	230	15	US-10-211-689-46
9	99.5	5.9	217	15	US-10-211-689-44
10	99.5	5.9	1057	12	US-10-282-122A-71880
11	98.5	5.8	782	16	US-10-437-963-148885
12	98.5	5.8	1493	16	US-10-437-963-132386
13	97.5	5.8	289	9	US-09-815-242-11940
14	97.5	5.8	289	12	US-10-282-122A-66521
15	97	5.7	485	15	US-10-369-493-17102

16	96	5.7	516	15	US-10-369-493-2886
17	95.5	5.7	797	9	US-09-815-242-50718
18	95.5	5.7	797	12	US-10-282-122A-43545
19	95.5	5.7	1057	12	US-10-282-122A-70305
20	94	5.6	155	9	US-09-738-626-5979
21	93.5	5.5	454	15	US-10-369-493-8439
22	93	5.5	498	9	US-09-886-156-59
23	93	5.5	498	9	US-09-886-150-59
24	93	5.5	498	10	US-09-886-149-59
25	93	5.5	498	10	US-09-886-159-59
26	93	5.5	498	14	US-10-326-090-59
27	92.5	5.5	233	15	US-10-211-689-32
28	92.5	5.5	239	15	US-10-211-689-36
29	92	5.4	445	9	US-09-815-242-5010
30	92	5.4	447	9	US-09-815-242-10908
31	92	5.4	447	12	US-10-282-122A-42511
32	91.5	5.4	444	12	US-10-282-122A-45519
33	91.5	5.4	685	16	US-10-437-963-118346
34	91.5	5.4	1222	16	US-10-408-765A-1650
35	91.5	5.4	1360	9	US-09-788-657-22
36	91.5	5.4	1360	10	US-09-912-697-16
37	91.5	5.4	1360	10	US-09-760-285-26
38	91.5	5.4	1360	14	US-10-270-839-37
39	91.5	5.4	1360	14	US-10-371-634-15
40	91.5	5.4	1360	14	US-10-348-074-40
41	91.5	5.4	1360	16	US-10-641-068-22
42	91.5	5.4	1360	16	US-10-408-765A-226
43	90.5	5.4	503	13	US-10-027-806-66
44	90.5	5.4	503	13	US-10-034-623-66
45	90.5	5.4	503	14	US-10-027-801-66

## ALIGNMENTS

## RESULT 1

US-10-627-476-314  
; Sequence 314, Application US/10627476  
; Publication No. US20040030116A1  
; GENERAL INFORMATION:  
; APPLICANT: Pompejus, Mark  
; APPLICANT: Kroger, Burkhard  
; APPLICANT: Schoder, Hartwig  
; APPLICANT: Zelder, Oskar  
; APPLICANT: Habernauer, Gregor  
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE  
; TITLE OF INVENTION: TRANSPORT  
; FILE REFERENCE: BGI-125CPCN  
; CURRENT APPLICATION NUMBER: US/10/627,476  
; PRIOR FILING DATE: 2003-07-25  
; PRIOR APPLICATION NUMBER: 09/602,787  
; PRIOR FILING DATE: 2000-06-23  
; PRIOR APPLICATION NUMBER: USSN 60/141031  
; PRIOR FILING DATE: 1999-06-25  
; PRIOR APPLICATION NUMBER: DE 19931454.3  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931478.0  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19931563.9  
; PRIOR FILING DATE: 1999-07-08  
; PRIOR APPLICATION NUMBER: DE 19932122.1  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932124.8  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932125.6  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932128.0  
; PRIOR FILING DATE: 1999-07-09  
; PRIOR APPLICATION NUMBER: DE 19932180.9  
; PRIOR FILING DATE: 1999-07-09  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 678

```
; SEQ ID NO 314
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-314

Query Match          7.0%; Score 118; DB 12; Length 708;
Best Local Similarity 25.5%; Pred. No. 0.0093;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLLALAGLPGDRWVDSTTPRLGVQKIMDSGE---HWAK-----P 75
DB 167 KFNIRCVRTPEELSALGPRDQGVTI----LGVKFAQQTGPTTRWEKDRRKLGDST 222
QY 76 YATGSRDFRKKTLROWVDNGFAVLN-ADNLNIATNSQLNE----YCLSDALQALRAYG 130
DB 223 YRLGSTNDKAVETTLRETQVKGAVVQAADNRRAANRAELRELERQYQASQELKVSWAQI 282
QY 131 TEGFEESLVVFLD-----EASKAVKARAEALQAAAMISVDLPGGEEF 171
DB 283 DVESADAAIAELDRLLLELNNTPEATLSARHEAAKQTLARVSDLLVAAQSEETVASMNL 342
QY 172 LLSPAGQNPLKKMWVEEVPFAPRSTVLYLGDTRGKHSLEPERIFEVLGLTFD 226
DB 343 KRAETELKRLSLPVAEVSEIEAREVEKFLANTRRVHA---ANVDEQITIALRED 394

RESULT 2
US-10-627-476-312
; Sequence 312, Application US/10627476
; Publication No. US20040030116A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Mark
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schoder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN MEMBRANE SYNTHESIS AND MEMBRANE
; TITLE OF INVENTION: TRANSPORT
; FILE REFERENCE: BGI-125CPCN
; CURRENT APPLICATION NUMBER: US/10/627,476
; CURRENT FILING DATE: 2003-07-25
; PRIOR APPLICATION NUMBER: 09/602,787
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: USSN 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931454.3
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931478.0
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931563.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932122.1
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932124.8
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932180.9
; PRIOR FILING DATE: 1999-07-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 678
; SEQ ID NO 312
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-627-476-312

Query Match          7.0%; Score 118; DB 12; Length 713;
Best Local Similarity 25.5%; Pred. No. 0.0094;
Matches 60; Conservative 27; Mismatches 104; Indels 44; Gaps 7;

QY 29 QRYNRSVAVTLLALAGLPGDRWVDSTTPRLGVQKIMDSGE---HWAK-----P 75
DB 570 KFNIRCVRTPEELSALGPRDQGVTI----LGVKFAQQTGPTTRWEKDRRKLGDST 625
QY 76 YATGSRDFRKKTLROWVDNGFAVLN-ADNLNIATNSQLNE----YCLSDALQALRAYG 130
DB 626 YRLGSTNDKAVETTLRETQVKGAVVQAADNRRAANRAELRELERQYQASQELKVSWAQI 685
QY 131 TEGFEESLVVFLD-----EASKAVKARAEALQAAAMISVDLPGGEEF 171
DB 686 DVESADAAIAELDRLLLELNNTPEATLSARHEAAKQTLARVSDLLVAAQSEETVASMNL 745
QY 172 LLSPAGQNPLKKMWVEEVPFAPRSTVLYLGDTRGKHSLEPERIFEVLGLTFD 226
DB 746 KRAETELKRLSLPVAEVSEIEAREVEKFLANTRRVHA---ANVDEQITIALRED 797
```

```
RESULT 4
US-10-211-689-34
; Sequence 34, Application US/10211689
; Publication No. US20030232347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khrantsov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/310,970
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 34
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-34

Query Match      6.2%; Score 105.5; DB 15; Length 229;
Best Local Similarity 23.1%; Pred. No. 0.034;
Matches 56; Conservative 39; Mismatches 114; Indels 33; Gaps 11;

QY 35 SAVTLLAAGLPGDRWDSTTPRLGVQKIMDW-SGEHWAKPYATGSRDFRKTLLRWV 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 AALLVTFLAGQAKVEQVETPEPELROQTQWQSQRW--ELALGRFWDYLR-----WV 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 94 DNGFAVLNADNINATNSOLNEYCLSDREALQALRAYTGPFEESLVVFLDEASKAVKAPA 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 59 QTLSEVQVEELUSSQTQELR--ALMDETMKELKAYSE--LEEQTTPVAETPRLSKEL 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 154 EALQAMISVDLPFGGEEFLLSPAGQ-NPLLKKWVEEFVPRFAP-----RSTVYLGD--- 204
```

```
Db 116 QAAQ-ARLGADMEDVRGRLVQYRGEVQAMLGQSTEEELRVRLASHLRKURKLLRDADDE 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 205 TRGKHSIFEREIFEVELGLTFDPHGRMPDLILHDEVRGWLFLMEAVK-----SKGPFDEE 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 175 EQAQIRLQAEAFQARKLSWFEP-----LVEDMQRWAGLVKVKVQAAVGTSAAPVPSD 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 RH 261
   |
Db 228 NH 229

RESULT 5
US-10-211-689-42
; Sequence 42, Application US/10211689
; Publication No. US20030232347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khrantsov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glennda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/310,970
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 42
; LENGTH: 236
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-42

Query Match      6.2%; Score 105.5; DB 15; Length 236;
```

```

Best Local Similarity 23.1%; Pred. No. 0.035;
Matches 56; Conservative 39; Mismatches 114; Indels 33; Gaps 11;

Qy 35 SAVTLALAGLPGDRWVDTTPRLGVOKIMDW-SGBHWAKPYATGSRDPRKKTLLRQWV 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 10 AALLVTEFLAGCAQAKVEQAVETEPPELRQQTQWQSGQRW--ELALGRFWDYLR-----WV 62

Qy 94 DNGFVAVNADNLNATNSQLNECLSDAQLALBAYGTGFEESLVVFLDEASKAVKARA 153
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 QTLSEQVQEEILLSSQVTELR--ALMDETMKELKAYKSE--LEEQLTPVABETRLRLSKEL 119

Qy 154 BALQAMISVDLPGLGEEFLSPAGQ-NPLLKKMVEEFVPRFAP-----RSTVLYLGD--- 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 QAAQ-ARLGADMEVGRGLVQYRGEVQAMLCQSTEEELRVRLASHLRLKELLRDADDLE 178

Qy 205 TRGHSLFEREIRFEIVGLTDPDHRMPDLILHDEVGRWLFLMEAVK-----SKGPDDEE 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 179 EQAQIRLQAEAFQARLKSNPEP-----LVEDMQRWAGLVEKVQAAVGTSAAPVPSD 231

Qy 260 RH 261
   |
Db 232 NH 233

```

RESULT 6

```

US-10-282-122A-70903
; Sequence 70903, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 03A4
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70903
; LENGTH: 1120
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-10-282-122A-70903

```

```

Query Match 6.1%; Score 103.5; DB 12; Length 1120;
Best Local Similarity 22.7%; Pred. No. 0.59;
Matches 61; Conservative 41; Mismatches 82; Indels 85; Gaps 14;

Qy 82 EDRFKKTLRWVNGFAVLNADNLN-----IATNSQLN--BY-CLSDAQLQALR 127
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 KEIEYVWRDKNDNAIVVCNMENIDPVGIHTGDSIVVAPSQTLSDVEYQMLRDVSLKVIR 632

Qy 128 AYGTGEG-----FEESLVVFLDEASKAVKARAELQ-----AAMISVDLPGG 168
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 633 ALGIEGCNVLADLPHSLNYYITIEVPRV-SRSSALASKATGYPIAKLAIAVGLTLD 691

Qy 169 EEFLLSP-----AGQNPLKKQWVEEFVPRF-----APRSTVLYLGDTR 206
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 692 E--MLNITGTSYAAFEPTLDYVISK-IPRFPDPKFKGERELCTQMKATGEVMAIGRT- 747

Qy 207 GKHSLFEREIPEEVGLTDFD-----PHGRMPDL-----ILHDEVGRWLFLMEAVKSG 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 748 -----YEESLLKATRSLEYGVHHLGLPENGESYELDYIKERIGHQDDERLFFIGBAIR--- 799

Qy 255 PFDEERHRSLOELFVTPSAGLIFVNCFFEN 283
   |
Db 800 -----RGTSLLEELHNMTKIDYFFLNFQON 823

```

RESULT 7

```

US-10-211-689-30
; Sequence 30, Application US/10211689
; Publication No. US20030232347A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John II
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Edinger, Shalomit R.
; APPLICANT: Gangolli, Esha A.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Khamstov, Nikolai V.
; APPLICANT: Lepley, Denise M.
; APPLICANT: MacDougall, John R.
; APPLICANT: Pena, Carol A.
; APPLICANT: Peyman, John A.
; APPLICANT: Patturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Smithson, Glennnda
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vernet, Corine A. M.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-416B
; CURRENT APPLICATION NUMBER: US/10/211,689
; CURRENT FILING DATE: 2003-01-21
; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13

```

```

: PRIOR APPLICATION NUMBER: 60/381,030
: PRIOR FILING DATE: 2002-05-16
: PRIOR APPLICATION NUMBER: 60/323,944
: PRIOR FILING DATE: 2001-09-21
: Remaining Prior Application data removed - See File Wrapper or PALM.
: NUMBER OF SEQ ID NOS: 132
: SOFTWARE: CuraSeqList version 0.1
: SEQ ID NO 30
: LENGTH: 194
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-211-689-30

Query Match      6.1%; Score 103; DB 15; Length 194;
Best Local Similarity 21.9%; Pred. No. 0.048;
Matches 51; Conservative 37; Mismatches 95; Indels 50; Gaps 10;

QY      35  SAVTLLAAGLAKPCDRWVDSITPLRGVQKINDW-SGEHWAKPYATGSGEDPRKKTILQWV 93
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6  AALLVTFLAGCQAKVEQAVETEPPEELRQQTWQSGQRW--ELALGRFWDYLR----WV 58
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      94  DNGFAVLNADNLNLTATNSQLNEYCLSDAQLALRAYGTGTFEESLVVFLDEASKAVKARA 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      59  QTLSEQVQEILLSSQVTELRL--ALMDETMKELKAYKSE-LEEQLTPVAETRLARKSEL 115
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      154  EALQAAAMISVDLPGEERFLLSPAGCNPLLKQWVEEFVFRAPSTRVYILGDTCKHSLFE 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      116  QAAQARL-----GADEV---KQVAEVRAKLEE-----QAQQLRLQ 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY      214  RETPEEVILGTFDPHGRMPDLILHDEVRGWLFIMEAVK-----SKGPFDBERH 261
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      149  AEAQFARLKSWFEP-----LVEDMQROWAGLVEKQAAVGT'SAAAPVPSDNH 194
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 8  
US-10-211-689-46  
; Sequence 46, Application US/10211689  
; Publication No. US20030232347A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, John II  
; APPLICANT: Anderson, David W.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Gorman, Linda  
; APPLICANT: Guo, Xiaojia (Sasha)  
; APPLICANT: Khamstov, Nikolai V.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Pena, Carol A.  
; APPLICANT: Peyman, John A.  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Voss, Edward Z.  
; APPLICANT: Zhong, Mei  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-416B  
; CURRENT APPLICATION NUMBER: US/10/211,689  
; CURRENT FILING DATE: 2003-01-21  
; PRIOR APPLICATION NUMBER: 60/311751  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/310,802  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/310,795  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/311,292  
; PRIOR FILING DATE: 2001-08-09

```

; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 46
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-211-689-46

Query Match      6.0%; Score 101; DB 15; Length 230;
Best Local Similarity 23.2%; Pred. No. 0.1;
Matches 53; Conservative 39; Mismatches 108; Indels 28; Gaps 10;

Qy 35 SAVTLLAGLKPGRWDSTTPRLGVQKIMDW--SGEHWAKPYATGSGRDFPKKTLRWV 93
Db 8 AALLVTFLACQAKVEQAVTEPEPRLQQTOWSGQRW--ELALGRFWDYLR-----WV 60
Qy 94 DNGFAVLNADNLNATINSQNLNEYCLSDAQLALRAYGTGEPGESLVVFLDEASKAVKARA 153
Db 61 QTLSEQVQEBLLSSQVTCQLR--ALMDETMKELKAYKSE-LEEQLTPVAETRLRLSKEL 117
Qy 154 EALQAAMISVDLPGBEEFLSPAGQ--NPLLKKMWVEEFPVPFAP-----RSTVLYLGD--- 204
Db 118 QAAQ-ARLQADMEDVRGLVQYRGVEVQAMLQSGSTEEFLVRVLASHLRLKRLLRDADDL 176
Qy 205 TRGKHSIFEREIEFEVVLGLTFDPHGRMPDDLILHDEVRGWLFLMEAVKS 252
Db 177 EQAQQLRLQAEAFQARLKSWFEP-----LVEDMQRWAGLVKEVKQA 217

```

RESULT 9  
US-10-211-689-44  
; Sequence 44, Application US/10211689  
; Publication No. US20030232347A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook, John II  
; APPLICANT: Anderson, David W.  
; APPLICANT: Boldog, Ferenc L.  
; APPLICANT: Burgess, Catherine E.  
; APPLICANT: Casman, Stacie J.  
; APPLICANT: Edinger, Shlomit R.  
; APPLICANT: Gangolli, Esha A.  
; APPLICANT: Gorman, Linda  
; APPLICANT: Guo, Xiaojia (Sasha)  
; APPLICANT: Khrantsov, Nikolai V.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: MacDougall, John R.  
; APPLICANT: Pena, Carol A.  
; APPLICANT: Peyman, John A.  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Rieger, Daniel K.  
; APPLICANT: Shimkets, Richard A.  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Spytek, Kimberly A.  
; APPLICANT: Vernet, Corine A. M.  
; APPLICANT: Voss, Edward Z.  
; APPLICANT: Zhong, Mei  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS  
; FILE REFERENCE: 21403-416B  
; CURRENT APPLICATION NUMBER: US/10/211.689  
; CURRENT FILING DATE: 2003-01-21

```

; PRIOR APPLICATION NUMBER: 60/311751
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/310,802
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/310,795
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/311,292
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 60/361,159
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: 60/373,050
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: 60/380,970
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/311,979
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/381,030
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/323,944
; PRIOR FILING DATE: 2001-09-21
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 132
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 44
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-211-689-44

Query Match          5.9%; Score 99.5; DB 15; Length 217;
Best Local Similarity 23.4%; Pred. No. 0.13;
Matches 52; Conservative 35; Mismatches 102; Indels 33; Gaps 11;

Qy 55 TTPRLGVQKIMDW-SGEHWAKPYATGSRDPKTKTLROWVNGFAVLNADNINATNSQL 113
Db 11 TEPEPELRQQTQWQSGQRW--ELALGREFWDYLR-----WVQTLSEQVQEELSSQVTOEL 63

Qy 114 NEYCLSEALQALRAYGTGFEERSIVVFLDRASKAVKARAEALQAAISVDLPQGEFFLL 173
Db 64 R-ALMDTETMKELKAYKSE-JEEQTPVAETRAKLSKELQAAQ-ARLGADMEDVGRGLV 119

Qy 174 SPAGQ-NPLLKKVMVEEFVRFPAP-----RSTVLYLGD---TRGKHSLEFEREIFEVLGLT 224
Db 120 QYRGEVQALQSGSTEELRVLRLASHLRKLRLDRADDLEEQAAQIRLQAAFAOARLKSW 179

Qy 225 FDPGRMPDLILHDEVGRWLFMEAVK-----SKGPFDEERH 261
Db 180 FEP-----LVEDMQRWAGLVEKVQAAVGTSAAPVPSDNH 214

RESULT 10
US-10-282-122A-71880
; Sequence 71880, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 71880
; LENGTH: 1057
; TYPE: PRT
; ORGANISM: Staphylococcus haemolyticus
; US-10-282-122A-71880

Query Match          5.9%; Score 99.5; DB 12; Length 1057;
Best Local Similarity 21.9%; Pred. No. 1.4;
Matches 59; Conservative 42; Mismatches 83; Indels 85; Gaps 14;

Qy 82 EDRPKTKLRQWVNGFAVLNADNINL-----IATNSQLN--EY-CLSDALQALR 127
Db 214 KEIEYEVNRDKNDNAIVVCNMENIDPVGIHTGDSIVVAPSQTLSDEVYQMLRDVSLKVIR 273

Qy 128 AYGTGEG-----FEESLVFLDEASKAVKARAEALQ-----AAMISVDLPQG 168
Db 274 ALGTGEGCNVOLADPHSFNYIIIEVNPVRV-SRSALASKATGPIAKLAKIAVGLTLD 332

Qy 169 EEFLLSP-----AGQNPLLKKVMVEEFVRPF-----APRSTVLYLGDTR 206
Db 333 E--MLNPVTGTSYAAFEPTLDYVISK-IPRPFDFKFKGERELGTQMKATGEVMAIGRT- 388

Qy 207 GKHSLEFEREIFEVLGLTDF-----PHGRMPDL-----ILHDEVGRWLFMEAVKSKG 254
Db 389 -----YERSLLKAIRSLLEYGVHGLPNGESFDLDYIKERISHQDDERLFFIGEAIR--- 440

Qy 255 PFDERHRSLQELFVTPSAGLIFVNCREN 283
Db 441 -----RGTTLEIHNMTOIDYFFLNKFN 464

RESULT 11
US-10-437-963-148885
; Sequence 148885, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 148885
; LENGTH: 782

```

```

; TYPE: PR1T
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_49272C.1.pep
US-10-437-963-148885

```

Query Match 5.8%; Score 98.5; DB 16; Length 782;  
Best Local Similarity 22.2%; Pred. No. 1.1;  
Matches 83; Conservative 37; Mismatches 125; Indels 129; Gaps 20;

Qy	15	DTARALLKRFQFDQRYNVRSVAVTLLALAGLKPGDRWVDS	TTPRLGVQKIMDSG-EHWA	73
Db	142	DTA-----VRFXFDYMMMMIRPVAVMLAVDLTLEGRASIDRN-	-RLGTSRLNIGGFQHF	195

Qy	74	KPATGTSREDFRK- : : : : : : : : : : : :	TLRQWVD- : : : : : : : : : : : :	NGFA- : : : : : : : : : : : :	98
Db	196	YKYPL-----NFRKNSNPNDPSFATTFVTITWEDQPQAGSDGIAFVLSSTKNLHNS			251

Qy	99	-----VLNADN-----	-----LN-----	-----IATNSQLNEYCLSD	120
Db	252	LGQYLGUFNASWTNSQNIILAIELDTFMNPLDMDMDNHVGIDVNSLHSTAGETST			311

CY	121	EALQALRAYGTGFEESLVVFLEDEASKVAKARAEALQAAMISVDLPGGEEFLLSPAGNP	180
D <sub>b</sub>	312	GGFQLRLR--ANGRSPIQLWLDYDGHKAHQLN-----VTGLP-----YSKPEYP	355

QY		181	LLKKMVEEFVPRAPSRVLYLGDTGRGKHSLFERIEFEEVLGLTFDPHGMPDL-----IL	236
Db		356	LUSIVN--LSLLPSSS---YIGFSASVNPKTRHF-----ILGWSFKENGVPPLPSVPVT	408

QY	237	HDEVRGW---	LFLMEAVKSKGPFDEERHRSLOEL--	FVTPSAGLIFVNCFNRESMRQWL	291
Db	409	DEPTYGMGNGFNFA	PPPPPLNTHQVHKH-SLQILLPI	MTSVILLLVAF-	457

QY	292	PBLAWETEANVAED	305
		:	
Dd	458	--LGWRKAGPOED	469

```

RESULT 12
US-10-437-963-132386
; Sequence 132386, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 132386
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1493)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3435C.1.pep
US-10-437-963-132386

```

Query Match 5.8%; Score 98.5; DB 16; Length 1493;  
Best Local Similarity 21.0%; Pred. NO. 3;  
Matches 78; Conservative 41; Mismatches 126; Indels 127; Gaps 17;

QY	3	SSDGIDGTVASID-----TARALKRFGFDAQRYNVRSVAITLLALAGLKPGRW-----	51
Db	542	ADDGAQGTVATLDGGGGXCAHTSRRGFKAMRRELLAAVPTHEAARKA--RWSEVKLT	598
QY	52	-----VDSTTPRLGVQKIMDWSGEHWAKPYATGSREDFRKKTLRQWVUNGFA	98
Db	599	FQSDHPTAPGMKLOPSUPIIGVTGHTWPLGHVELPMTFGDSTNFRTER-----NDPD	652
QY	99	VLNADNLNIATNSQUNEYCL-----SDEALQALRAYGTEGFEESLVVFLD-----	143
Db	653	VA---DLNLFPYNVLGRPTLVKFMAATHYAVLQMKMQGPAG--PITVFGDVKVALTCAE	706
QY	144	-----EASKAVKARAEFA-IQAAMI5VDLPGGEEFLILSPAQCNPLLKCM	185
Db	707	QCADNLAVATEPQAPAEASRKKRLTSADEBEGALVSFLRANSDFVAMKLSMPRVIKEV	766
QY	186	VEEFVPRFAPRSTVLYLGDTRGKSLPEREPIEEVLGHTPDPHGRMPDILIHDEVRGWI--	244
Db	767	IEH--RFAVRQKV-----RRQALERQAF-----FREEV-AWLL	796
QY	245	---FLMEAVKSGKPFDEERHRSLOELFVTPSAGLIFVNCENRESMRQWLPELAEATEAW	301
Db	797	EADFIREVHPW-----LANLVAPKANGKLRMCIDYTD-LINKACP-----	837
QY	302	VAEDPDHLIHLN	313
Db	838	--KDPFHLPHID	847

```

RESULT 13
US-09-815-242-11940
; Sequence 11940, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11940
; LENGTH: 289
; TYPE: PR1
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-11940

```

Query Match 5.8%; Score 97.5; DB 9; Length 289;  
Best Local Similarity 24.0%; Pred. No. 0.32;  
Matches 59; Conservative 32; Mismatches 90; Indels 6



Mon Oct 4 09:05:16 2004

us-10-668-047-4.rapb

Page 9

Search completed: October 1, 2004, 16:28:47  
Job time : 71.0717 secs

Blank